

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:27:25 ; Search time 82.85 Seconds  
(without alignments) 57.648 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGCGKGANARQDKS.....GDIAQALGSGSIAIGDKIV 43

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	43	18	AAW32274
2	209	100.0	43	22	AAE00703
3	209	100.0	2123	22	AAE00701
4	190	90.9	40	22	AAE00705
5	121	57.9	24	18	AAW32276
6	121	57.9	24	22	AAE00702
7	90	43.1	1992	17	AAW04505
8	90	43.1	1992	22	AAE09133
9	90	43.1	1992	22	AAE09137
10	90	43.1	2047	22	AAE09134
11	90	43.1	2053	22	AAE09135

12	90	43.1	2314	22	AAE09136
13	68	32.5	1778	22	AAE09137
14	67	32.1	2042	19	AAE09139
15	63	30.1	261	20	AAE09148
16	61	29.2	2039	19	AAE09132
17	60.5	28.9	608	22	AAE09128
18	60	28.7	278	15	AAE09140
19	60	28.7	278	17	AAE09142
20	60	28.7	278	18	AAE09144
21	60	28.7	278	22	AAE09146
22	60	28.7	278	21	AAE09148
23	60	28.7	278	21	AAE09149
24	60	28.7	278	21	AAE09150
25	60	28.7	278	21	AAE09151
26	58.5	28.0	449	20	AAE09152
27	56	26.8	591	22	AAE09153
28	55.5	26.6	240	22	AAE09154
29	55.5	26.6	919	22	AAE09155
30	55	26.3	193	20	AAE09156
31	55	26.3	252	21	AAE09157
32	55	26.3	889	19	AAE09158
33	55	26.3	1114	22	AAE09159
34	55	26.3	1173	22	AAE09160
35	54.5	26.1	166	21	AAE09161
36	54.5	26.1	187	22	AAE09162
37	54.5	26.1	467	20	AAE09163
38	54	25.8	456	22	AAE09164
39	54	25.8	456	22	AAE09165
40	53.5	25.6	700	22	AAE09166
41	53.5	25.6	704	21	AAE09167
42	53.5	25.6	704	21	AAE09168
43	53.5	25.6	704	22	AAE09169
44	53.5	25.6	704	22	AAE09170
45	53.5	25.6	1438	22	AAE09171

# ALIGNMENTS

RESULT 1	AAW32274	standard: peptide; 43 AA.
XX	AAW32274:	
AC	08-MAY-1998	(first entry)
XX		
DE	M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.	
XX		
KW	Outer membrane protein-106; OMP106; vaccine; immune response;	
KW	Cytotoxic antibody; Moraxella catarrhalis.	
OS	Moraxella catarrhalis.	
XX		
PN	WO9741731-A1.	
XX		
PD	13-NOV-1997.	
XX		
PF	28-APR-1997; 97WO-US07679.	
XX		
PR	03-MAY-1996; 96US-0642712.	
XX		
PA	(ANTE-) ANTEX BIOLOGICS INC.	
XX		
PI	Plosilla L, Tucker K;	
XX		
DR	WPI; 1997-558601/51.	
XX		
PT	Outer membrane protein, OMP106, of Moraxella catarrhalis - used in	
XX	vaccines for producing immune responses against M. catarrhalis	
PS	Claim 9; Page 23; 78pp; English.	
XX		

M. catarrhalis les  
Escherichia coli p  
Haemophilus paraga  
Mouse clone 65 pro  
Haemophilus paraga  
Drosophila melanog  
TARA-binding prote  
TARA-binding prote  
Drosophila TARA-bi  
TARA-binding prote  
Drosophila melanog  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Tobacco geranyl ge  
Novel human diago  
Human TE anti-idio  
Novel human diago  
Protein involved i  
Amino acid sequenc  
M. catarrhalis str  
Novel human diago  
Novel human diago  
Zea mays protein f  
Drosophila melanog  
A. thaliana GGPOR  
Human clone 65 pro  
Drosophila melanog  
Amino acid sequenc  
Haemophilus influe  
E. coli proliferat  
E. coli cellular p  
Salmonella typhi c  
Novel human diago

CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)  
CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella  
CC catarrhalis, an haemagglutinating cultivar. The peptide fragment can  
CC specifically bind to an antibody that binds the OMP106 polypeptide. The  
CC antibody is a cytotoxic antibody which mediates complement killing of  
CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be  
CC used in vaccines and antigenic compositions. They can also be used for  
CC producing an immune response in an animal against M. catarrhalis.  
XX  
SQ Sequence 43 AA;

Query Match 100.0%; Score 209; DB 18; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.1e-19;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 IGISEADGKGANARDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 1 IgiSeadgkgganarYdksiaIgdIagalsgsiaIgdnkiv 43

## RESULT 2

ID AAE00703 standard; peptide; 43 AA.

AC AAE00703;

DT 02-JUL-2001 (first entry)

DE N-terminal #1 of M. catarrhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;  
KW bacterial infection; immunogen; cytotoxic; antibiotic;  
KW passive immunisation.

OS Moraxella catarrhalis.

XX US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarrhalis outer  
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial  
PT infections and as vaccine against Moraxella catarrhalis infection of  
PT mammals -

PS Claim 2; Column 31; 49pp; English.

CC The present sequence is N-terminal of haemagglutinating  
CC Moraxella catarrhalis outer membrane protein-106 (OMP106). The OMP106 is  
CC used as a therapeutic and prophylactic vaccine against M. catarrhalis  
CC infections of mammals. It is used for diagnosis of bacterial infections  
CC and as reagents for clinical or medical diagnosis of M. catarrhalis  
CC infections and for scientific research on the properties of  
CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also  
CC used as a probe to identify the presence of M. catarrhalis in biological  
CC specimens and to identify other bacteria that encode a polypeptide  
CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used  
CC as ligands to detect antibodies elicited in response to M. catarrhalis  
CC infections and also as immunogens for inducing M. catarrhalis-specific  
CC antibodies which are useful in immunoassays to detect M. catarrhalis in  
CC biological specimens. Cytotoxic antibodies are useful in passive  
CC immunisations against M. catarrhalis.

XX  
SQ Sequence 43 AA;

Query Match 100.0%; Score 209; DB 22; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.1e-19;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 IGISEADGKGANARDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 1 IgiSeadgkgganarYdksiaIgdIagalsgsiaIgdnkiv 43

## RESULT 3

ID AAE00701 standard; Protein; 2123 AA.

AC AAE00701;

DT 02-JUL-2001 (first entry)

DE Moraxella catarrhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;  
KW bacterial infection; immunogen; cytotoxic; antibiotic;  
KW passive immunisation.

OS Moraxella catarrhalis.

XX US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarrhalis outer  
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial  
PT infections and as vaccine against Moraxella catarrhalis infection of  
PT mammals -

PS Claim 7; Column 53-64; 49pp; English.

CC The present sequence is haemagglutinating Moraxella catarrhalis outer  
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic  
CC and prophylactic vaccine against M. catarrhalis infections of mammals.  
CC It is used for diagnosis of bacterial infections and as reagents for  
CC clinical or medical diagnosis of M. catarrhalis infections and for  
CC scientific research on the properties of pathogenicity, virulence and  
CC infectivity of M. catarrhalis. It is also used as a probe to identify  
CC the presence of M. catarrhalis in biological specimens and to identify  
CC other bacteria that encode a polypeptide related to M. catarrhalis  
CC OMP106. OMP106-derived polypeptides are used as ligands to detect  
CC antibodies elicited in response to M. catarrhalis infections and also  
CC as immunogens for inducing M. catarrhalis-specific antibodies which are  
CC useful in immunoassays to detect M. catarrhalis in biological specimens.  
CC Cytotoxic antibodies are useful in passive immunisations against  
CC M. catarrhalis.

SQ Sequence 2123 AA;

Query Match 100.0%; Score 209; DB 22; Length 2123;  
Best Local Similarity 100.0%; Pred. No. 9e-18;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGGANARCKSTAIGDIAQALGSQSIATGDKNTIV 43  
 DB 69 igiseadgkkganargdksiaigdiagalsygsiaigdnkiv 111

## RESULT 4

AAE00705 ID AAE00705 standard; peptide: 40 AA.

AC AAE00705;

DT 02-JUL-2001 (first entry)

DE N-terminal #2 of M. catarrhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antituberc;

OS Moraxella catarrhalis.

PN US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman WF;

DR WPI; 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarrhalis outer

PT membrane protein-106 polypeptide, useful for diagnosis of bacterial

PT infections and as vaccine against Moraxella catarrhalis infection of

PT mammals -

PS Example: Column 31; 49pp; English.

CC The present sequence is N-terminal of haemagglutinating

CC Moraxella catarrhalis outer membrane protein-106 (OMP106). The OMP106 is

CC used as a therapeutic and prophylactic vaccine against M. catarrhalis

CC infections of mammals. It is used for diagnosis of bacterial infections

CC and as reagents for clinical or medical diagnosis of M. catarrhalis

CC infections and for scientific research on the properties of

CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also

CC used as a probe to identify the presence of M. catarrhalis in biological

CC specimens and to identify other bacteria that encode a polypeptide

CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used

CC as ligands to detect antibodies elicited in response to M. catarrhalis

CC infections and also as immunogens for inducing M. catarrhalis-specific

CC antibodies which are useful in immunoassays to detect M. catarrhalis in

CC biological specimens. Cytotoxic antibodies are useful in passive

CC immunisations against M. catarrhalis.

CC Sequence 40 AA;

SO

Query Match 90.9%; Score 190; DB 22; Length 40;

Best Local Similarity 100.0%; Pred. No. 2.6e-17;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGGANARCKSTAIGDIAQALGSQSIATGDKNTIV 39

DB 2 igiseadgkkganargdksiaigdiagalsygsiaigdnkiv 40

RESULT 5

AAW32276

ID AAW32276 standard; peptide: 24 AA.

AC AAW32276;

DT 08-MAY-1998 (first entry)

DE M. catarrhalis outer membrane protein (OMP)-106 partial peptide fragment.

KW Outer membrane protein-106; OMP106; vaccine; immune response;

KM cytotoxic antibody; Moraxella catarrhalis; primer; probe.

OS Moraxella catarrhalis.

PN WO9741731-A1.

PD 13-NOV-1997.

PF 28-APR-1997; 97WO-US07679.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Plosila L, Tucker K;

DR WPI; 1997-558601/51.

DR N-FSDP; AAT86522.

PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in

PT vaccines for producing immune responses against M. catarrhalis

PS Disclosure; Page 58; 78pp; English.

CC This is a partial sequence of an amino terminal peptide fragment of a

CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an

CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating

CC cultivar. The encoding DNA can be used as a 5' primer for PCR

CC amplification of a full length OMP106 DNA. The DNA fragment can also be

CC used as a probe for screening M. catarrhalis genomic libraries for OMP106

CC polypeptide coding sequences. The encoded peptide fragment can

CC specifically bind to an antibody that binds the OMP106 polypeptide. The

CC antibody is a cytotoxic antibody which mediates complement killing of

CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be

CC used in vaccines and antigenic compositions. They can also be used for

CC producing an immune response in an animal against M. catarrhalis.

CC Sequence 24 AA;

SO

Query Match 57.9%; Score 121; DB 18; Length 24;

Best Local Similarity 100.0%; Pred. No. 8.8e-09;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 EADGKGGANARCKSTAIGDIAQ 28

DB 1 aadgkkganargdksiaigdiag 24

RESULT 6

AAE00702 ID AAE00702 standard; peptide: 24 AA.

AC AAE00702;

DT 02-JUL-2001 (first entry)

DE Internal fragment of N-terminal outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antituberc;

KW passive immunisation.

OS Moraxella catarrhalis.

XX US6214981-B1.  
 XX 10-APR-2001.  
 XX  
 XX 12-NOV-1997; 97US-0968685.  
 XX  
 XX 03-MAY-1996; 96US-0642712.  
 XX  
 XX (ANTE-) ANTEX BIOLOGICS INC.  
 XX  
 XX Tucker K, Plosila L, Tillman UF;  
 XX WPI: 2001-281002/29.  
 XX N-PSDB: AAD04030.  
 XX  
 XX Novel nucleotide sequences encoding Moraxella catarrhalis outer  
 XX membrane protein-106 polypeptide, useful for diagnosis of bacterial  
 XX infections and as vaccine against Moraxella catarrhalis infection of  
 XX mammals -

Example; Column 43-44; 49pp; English.

The present sequence is the internal fragment of N-terminal  
 haemagglutinating Moraxella catarrhalis outer membrane protein-106  
 (OMP106). This sequence is used to design a probe and a 5' PCR primer.  
 The OMP106 is used as a therapeutic and prophylactic vaccine against  
 M. catarrhalis infections of mammals. It is used for diagnosis of  
 bacterial infections and as reagents for clinical or medical diagnosis  
 of M. catarrhalis infections and for scientific research on the  
 properties of pathogenicity, virulence and infectivity of M. catarrhalis.  
 It is also used as a probe to identify the presence of M. catarrhalis in  
 biological specimens and to identify other bacteria that encode a  
 polypeptide related to M. catarrhalis OMP106. OMP106-derived polypeptides  
 are used as ligands to detect antibodies elicited in response to  
 M. catarrhalis infections and also as immunogens for inducing  
 M. catarrhalis-specific antibodies which are useful in immunoassays to  
 detect M. catarrhalis in biological specimens. Cytotoxic antibodies are  
 useful in passive immunisations against M. catarrhalis.

Sequence 24 AA;

Query Match 57.9%; Score 121; DB 22; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 8; 8e-09;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 EADGKGGANARGDKSTAIQDIAQ 28  
 |||  
 Db 1 eadgkgyganargksiaigdiaq 24

RESULT 7

AAW04505  
 ID AAW04505 standard; Protein; 1992 AA.  
 XX  
 XX AAW04505;

XX 25-JAN-1997 (first entry)

XX Moraxella 200 kDa outer membrane protein.

XX Outer membrane protein; OMP; immunogen; vaccine; otitis media;  
 XX diagnosis.

XX Mycobacterium catarrhalis strain 4223.

XX WO9634960-A1.

XX 07-NOV-1996.

XX 29-APR-1996; 96WO-CA00264.

XX

PR 26-MAR-1996; 96US-0621944.  
 PR 01-MAY-1995; 95US-0431718.  
 PR 07-JUN-1995; 95US-0478370.  
 XX

XX (CONN-) CONNAUGHT LAB LTD.

XX Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;

XX WPI: 1996-506162/50.

XX N-PSDB: AAT38740.

XX Moraxella outer membrane protein - useful as immunogen in protective

XX vaccine and for diagnosis

XX Claim 14; Fig 6; 109pp; English.

XX An approx. 200 kDa outer membrane protein (AAW04505) can be  
 CC isolated from Moraxella catarrhalis otitis media strain 4223  
 CC by electroporation, or expressed from a gene (see also AAT38740)  
 CC obtd. from a strain 4223 genomic library. Natural or recombinant  
 CC outer membrane protein is useful as an immunogen to protect  
 CC against infection by Moraxella, esp. M. catarrhalis. It can  
 CC also be used to detect antibodies, esp. for differential diagnosis  
 CC between bacteria that cause similar symptoms, and also useful as  
 CC a carrier for other antigens and used to raise antitumour  
 CC antibodies for conjugation to therapeutic agents.

SO Sequence 1992 AA;

Query Match 43.1%; Score 90; DB 17; Length 1992;  
 Best Local Similarity 52.3%; Pred. No. 0.012;  
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSTAIQDIAQALGSGSTAIQDNKIV 43  
 |||  
 Db 1718 akadgaavairgrtqgngsaiqdnagatgdsiaigtgnv 1761

RESULT 8

AAW69133  
 ID AAW69133 standard; Protein; 1992 AA.

XX AAW69133;

XX 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;  
 XX otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI: 2001-159722/16.

XX N-PSDB: AAF59100, AAF59101.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
 XX useful in protective vaccines and for diagnosis -



PS Example 3; Fig 2A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* strain 4223 lambdaDEMBL3 clone 200kDa protein, which is

CC used in the exemplification of the present invention.

XX

SO Sequence 1992 AA;

Query Match 43.1%; Score 90; DB 22; Length 1992;

Best Local Similarity 52.3%; Pred. No. 0.012;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGKANAR---GDKSIAIGDIAQLGSGSIAIGDNKIV 43

DB 1718 akadgaavaiqrqtqngsialqdnagatgdsialgtgnv 1761

RESULT 9

AAB69137

ID AAB69137 standard; Protein: 1992 AA.

XX

AC AAB69137;

XX

DT 24-APR-2001 (first entry)

XX

DE M. catarrhalis M56 200kDa protein in PK3348 SEQ ID NO:13.

XX

KW Moraxella catarrhalis strain O8; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KM otitis media; detection.

XX

OS Moraxella catarrhalis.

XX

PN WO200107619-A1.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-CA00870.

XX

PR 27-JUL-1999; 99US-0361619.

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX

PT WPI: 2001-159722/16.

XX

DR N-PSDB; AAF59106.

XX

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

XX

PS Claim 1; Fig 8A-V; 247pp; English.

XX

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* M56/200kDa protein in PK3348, which is given in the

CC exemplification of the present invention.

XX

SO Sequence 1992 AA;

Query Match 43.1%; Score 90; DB 22; Length 1992;

Best Local Similarity 52.3%; Pred. No. 0.012;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGKANAR---GDKSIAIGDIAQLGSGSIAIGDNKIV 43

DB 1718 akadgaavaiqrqtqngsialqdnagatgdsialgtgnv 1761

RESULT 10

AAB69134

ID AAB69134 standard; Protein: 2047 AA.

XX

AC AAB69134;

XX

DT 24-APR-2001 (first entry)

XX

DE M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.

XX

KW Moraxella catarrhalis strain 4223; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KM otitis media; detection.

XX

OS Moraxella catarrhalis.

XX

PN WO200107619-A1.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-CA00870.

XX

PR 27-JUL-1999; 99US-0361619.

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX

PT WPI: 2001-159722/16.

XX

DR N-PSDB; AAF59102, AAF59103.

XX

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

XX

PS Claim 1; Fig 3A-W; 247pp; English.

XX

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* strain 4223 genomic 200kDa protein, which is given in the

CC exemplification of the present invention.

SQ Sequence 2047 AA;

Query Match 43.1%; Score 90; DB 22; Length 2047;

Best Local Similarity 52.3%; Pred. No. 0.013;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGKGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 1773 akadgaavaiaigrtgagngsialgdnagatgdsiaigtgnv 1816

RESULT 11

AAB69135 ID AAB69135 standard; Protein; 2053 AA.

XX AAB69135;

XX 24-APR-2001 (first entry)

DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.

KM Moraxella catarrhalis strain Q8; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; Immunogenic; Infection;

XX otitis media; detection.

XX Moraxella catarrhalis.

XX MO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59104.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis  
XX Claim 1; Fig 4A-V; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis strain Q8 200kDa protein, which is given in the  
CC exemplification of the present invention.

SQ Sequence 2053 AA;

Query Match 43.1%; Score 90; DB 22; Length 2053;

Best Local Similarity 52.3%; Pred. No. 0.013;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGKGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 1773 akadgaavaiaigrtgagngsialgdnagatgdsiaigtgnv 1816

Db 1779 akadgaavaiaigrtgagngsialgdnagatgdsiaigtgnv 1822

RESULT 12

AAB69136 ID AAB69136 standard; Protein; 2314 AA.

XX AAB69136;

XX 24-APR-2001 (first entry)

DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.

KM Moraxella catarrhalis strain Q8; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; Immunogenic; Infection;

XX otitis media; detection.

XX Moraxella catarrhalis.

XX MO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59105.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
XX useful in protective vaccines and for diagnosis  
XX Claim 1; Fig 5A-Y; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)  
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
CC can be used in vaccines. (II), and its truncated versions, are used as  
CC immunogenic compositions and vaccines to protect against M. catarrhalis  
CC infections, particularly otitis media in humans. (II) is also used as  
CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
CC generate Ab. (I) are used for recombinant production of (II) and its  
CC fragments are used as probes for identifying/cloning 200 kDa protein  
CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
CC (I) makes possible production of large amount of recombinant immunogens.  
CC Expression of truncated versions of (II) reduces toxicity of the protein  
CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis les1 200kDa protein, which is given in the exemplification  
CC of the present invention.

SQ Sequence 2314 AA;

Query Match 43.1%; Score 90; DB 22; Length 2314;

Best Local Similarity 52.3%; Pred. No. 0.014;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGKGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 2040 akadgaavaiaigrtgagngsialgdnagatgdsiaigtgnv 2083

RESULT 13

AAB52677 ID AAB52677 standard; Protein; 1778 AA.

XX AAB52677;

QY 4 SEADGKGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43  
Db 1773 akadgaavaiaigrtgagngsialgdnagatgdsiaigtgnv 1816

DT 11-FEB-2002 (first entry)  
 XX  
 DE Escherichia coli polypeptide SEQ ID NO 749.  
 XX  
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
 KW systemic infection; non-diarrhoeal infection; septicemia;  
 KW pyelonephritis; antibiotic resistance.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200166572-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 12-MAR-2001; 2001WO-EP03445.  
 XX  
 PR 10-MAR-2000; 2000FR-0003145.  
 XX  
 PR 02-FEB-2001; 2001FR-0001449.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;  
 XX  
 DR WPI: 2001-550253/61.  
 XX  
 PS A library of DNA fragments of Escherichia coli strains for the  
 PT phylogenetic determination of a given strain comprises polynucleotides of  
 PT nature B2/D+ A- -  
 XX  
 PS Example 6; Fig 6; 646pp: English.  
 XX  
 CC The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)  
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature  
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as  
 CC septicemia, pyelonephritis and meningitis this is particularly  
 CC advantageous as bacterial resistance is increasing with the more  
 CC frequent use of broad spectrum antibiotics.  
 CC  
 CC Sequence 1778 AA;  
 XX  
 SQ

Query Match 32.5%; Score 68; DB 22; Length 1778;  
 Best Local Similarity 53.8%; Pred. No. 6.8;  
 Matches 14; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
 OY 15 ARGDKSIAGDIAQALGSQSIAGDN 40  
 DB 170 adgkxtialgnkayelmsialgdn 195  
 |||:||||:|:| |||:||||  
 RESULT 14  
 AAM56319  
 ID AAM56319 standard; Protein; 2042 AA.  
 XX  
 AC AAM56319;  
 XX  
 DT 19-AUG-1998 (first entry)  
 XX  
 DE Haemophilus paragallinarum antigenic protein#1.  
 XX  
 KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;  
 KW vaccine; chicken infectious coryza; CIC; fowl.  
 XX  
 OS Haemophilus paragallinarum.  
 XX

FH Key Location/Qualifiers  
 FT Peptide 1..70  
 FT Protein /label= signal  
 FT 71..2042  
 FT /note= "antigenic protein"  
 PN WO9812331-A1.  
 XX  
 PD 26-MAR-1998.  
 XX  
 PF 12-SEP-1997; 97WO-JP03222.  
 XX  
 PR 19-SEP-1996; 96JP-0271408.  
 XX  
 PA (KAGA ) ZH KAGAKU & KESSSEI RYOHO KENKYUSHO.  
 PA (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.  
 PI Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;  
 XX  
 DR WFI: 1998-230318/20.  
 DR N-FSDB; AAV22834.  
 XX  
 PT Antigenic polypeptide from Haemophilus paragallinarum induces HI  
 PT antibody production - and is useful for diagnosis of and preparation  
 PT of vaccines for chicken infectious coryza  
 XX  
 PS Claim 1; Page 51-69; 108pp; Japanese.  
 XX  
 CC The present sequence represents an antigenic protein derived from  
 CC Haemophilus paragallinarum strain A-221. The antigenic protein  
 CC stimulates the production of HI antibodies in fowl. The protein  
 CC and DNA coding for it can be used in the preparation of vaccines  
 CC for the prevention of chicken infectious coryza (CIC). The protein  
 CC and its antibodies can be used in the diagnosis and treatment of CIC.  
 CC  
 CC Sequence 2042 AA;  
 XX  
 SQ

Query Match 32.1%; Score 67; DB 19; Length 2042;  
 Best Local Similarity 45.2%; Pred. No. 11;  
 Matches 14; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
 OY 9 GKGGMARGDKSIAGDIAQALGSQSIAGD 39  
 DB 322 gkigaialgrtiaaemstavgslafalaad 352  
 ||| ||| |||:||||:|:| |||:||||  
 RESULT 15  
 AAY13486  
 ID AAY13486 standard; Protein; 261 AA.  
 XX  
 AC AAY13486;  
 XX  
 DT 26-JUL-1999 (first entry)  
 XX  
 DE Mouse clone 65 protein.  
 XX  
 KW Human; murine; clone 65; clone 320; Wnt; arteriosclerosis; malignancy;  
 KW melanoma; cancer; breast; ovary; colon; tumor; cardiac; renal;  
 KW inflammation; angiogenesis; immunological disorder.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9921999-A2.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 29-OCT-1998; 98WO-US22992.  
 XX  
 PR 04-FEB-1998; 98US-0073612.  
 PR 29-OCT-1997; 97US-0063704.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX

XX Levine AJ, Pennica D;  
 XX  
 DR WPI: 1999-326705/27.  
 DR N-PSDB; AAX55608, AAX55609.

PT Human and murine genes induced by Wnt and derived polypeptides

PS Claim 6; Fig 1A-B; 114pp; English.

XX  
 CC The invention relates to human and murine cDNA clones 65 and 320,  
 CC encoding polypeptides that are induced by Wnt. Clone 65 and 320  
 CC polypeptides, and their antagonists, are used to treat disorders  
 CC associated with these polypeptides, e.g. arteriosclerosis, malignancies  
 CC (particularly melanoma or cancer of the breast, ovary and colon), but  
 CC also benign tumors, cardiac, renal, catabolic or many other types of  
 CC disease. Inflammation, angiogenesis and immunological disorders, more  
 CC generally to induce death of Wnt-induced cells. Antibodies against the  
 CC polypeptides are used as therapeutic antagonists and as diagnostic  
 CC immunoassay or affinity-purification reagents. Clone 65 and 320 nucleic  
 CC acids are used as hybridization probes or primers (to detect related  
 CC sequences); for chromosome and gene mapping; to generate antisense  
 CC sequences; for recombinant polypeptide production and to generate  
 CC transgenic or 'knockout' animals (for development of, and screening for,  
 CC drugs). The polypeptides are used to raise (or purify) specific  
 CC antibodies; as immunoassay reagents and in drug screens.

XX  
 SQ Sequence 261 AA;

Query Match 30.1%; Score 63; DB 20; Length 261;

Best Local Similarity 38.9%; Pred. No. 3.3;

Matches 14; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

OY 2 GISEADGGKGGANARGPDKSIAIGDIAQALGSOSIAI 37

DB 37 gvsgrgraggaeqrqvkcvlygd--gavgktsivv 70

Search completed: July 30, 2002, 15:34:11  
 Job time: 406 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:30:55 ; Search time 33.83 Seconds  
(without alignments)  
31.046 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209  
Sequence: 1 IGISEADGKGANRGRDGS.....GDIAQALSGSIAIDNKIV 43

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfill1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	209	100.0	43	4	US-08-968-685A-1		Sequence 1, Appl
2	209	100.0	2123	4	US-08-968-685A-10		Sequence 10, Appl
3	190	90.9	40	4	US-08-968-685A-12		Sequence 12, Appl
4	121	57.9	24	4	US-08-968-685A-5		Sequence 5, Appl
5	121	57.9	24	4	US-08-968-685A-13		Sequence 13, Appl
6	121	57.9	24	4	US-08-968-685A-19		Sequence 19, Appl
7	90	43.1	2314	4	US-09-268-347-49		Sequence 49, Appl
8	89	42.6	2048	4	US-09-268-347-48		Sequence 48, Appl
9	60	28.7	278	1	US-08-188-582-9		Sequence 9, Appl
10	60	28.7	278	1	US-08-646-715-9		Sequence 9, Appl
11	55	26.3	379	4	US-09-180-827-12		Sequence 12, Appl
12	55	26.3	889	4	US-09-336-447A-15		Sequence 15, Appl
13	53.5	25.6	335	2	US-08-405-175A-6		Sequence 6, Appl
14	52.5	25.1	1160	3	US-08-808-559A-24		Sequence 24, Appl
15	52	24.9	513	3	US-08-173-508-6		Sequence 6, Appl
16	52	24.9	513	3	US-08-265-310-6		Sequence 6, Appl
17	52	24.9	513	3	US-08-951-742-6		Sequence 6, Appl
18	51	24.4	580	2	US-08-591-079-2		Sequence 2, Appl
19	50.5	24.2	126	3	US-07-765-830A-1		Sequence 1, Appl
20	50.5	24.2	126	3	US-07-765-830A-6		Sequence 6, Appl
21	50.5	24.2	126	3	US-07-765-830A-7		Sequence 7, Appl
22	50.5	24.2	768	2	US-08-222-617A-5		Sequence 5, Appl
23	50.5	24.2	3666	2	US-08-222-617A-12		Sequence 12, Appl
24	50.5	24.2	3727	2	US-08-222-617A-27		Sequence 27, Appl
25	50.5	24.2	3778	2	US-08-222-617A-2		Sequence 2, Appl
26	50	23.9	249	2	US-08-797-689-18		Sequence 18, Appl
27	50	23.9	1002	4	US-09-268-347-24		Sequence 24, Appl

28	50	23.9	1004	4	US-09-268-347-30	Sequence 30, Appl
29	49.5	23.7	430	1	US-08-535-237-2	Sequence 2, Appl
30	49.5	23.7	892	4	US-09-336-447A-5	Sequence 5, Appl
31	49	23.4	126	2	US-07-757-606B-1	Sequence 1, Appl
32	49	23.4	126	2	US-07-757-606B-6	Sequence 6, Appl
33	49	23.4	126	3	US-07-728-220C-20	Sequence 20, Appl
34	49	23.4	357	4	US-08-838-151A-20	Sequence 20, Appl
35	49	23.4	624	4	US-09-336-447A-7	Sequence 7, Appl
36	48.5	23.2	248	2	US-08-887-352B-22	Sequence 22, Appl
37	48.5	23.2	248	2	US-08-887-352B-23	Sequence 23, Appl
38	48.5	23.2	248	4	US-09-109-207C-22	Sequence 22, Appl
39	48.5	23.2	248	4	US-09-109-207C-23	Sequence 23, Appl
40	48.5	23.2	248	4	US-09-296-005-22	Sequence 22, Appl
41	48.5	23.2	248	4	US-09-296-005-23	Sequence 23, Appl
42	48.5	23.2	941	4	US-09-336-447A-9	Sequence 9, Appl
43	48	23.0	240	1	US-08-488-113B-148	Sequence 148, App
44	48	23.0	240	1	US-08-477-484B-148	Sequence 148, App
45	48	23.0	240	2	US-08-646-360-148	Sequence 148, App

#### ALIGNMENTS

RESULT 1  
US-08-968-685A-1  
Sequence 1, Application US/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:  
APPLICANT: PLOSILA, KENNETH  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: EDMONDS LLP  
ADDRESS: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-1

Query Match 100.0%; Score 209; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 3.6e-20;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 IGISEADGKGANRGRDGSIAIDIAQALSGSIAIDNKIV 43  
|||||

Db 1 IGISEADGKGANARGDKSIAIGDIAQALGSGSIAIGDNKIV 43

## RESULT 2

US-08-968-685A-10  
Sequence 10, Application US/08968685A  
Patent No. 6214981

## GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSILA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2123 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-10

Query Match 100.0%; Score 209; DB 4; Length 2123;  
Best Local Similarity 100.0%; Pred. No. 3.3e-18;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISEADGKGANARGDKSIAIGDIAQALGSGSIAIGDNKIV 43  
DB 69 IGISEADGKGANARGDKSIAIGDIAQALGSGSIAIGDNKIV 111

## RESULT 3

US-08-968-685A-12  
Sequence 12, Application US/08968685A  
Patent No. 6214981

## GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSILA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-12

Query Match 90.9%; Score 190; DB 4; Length 40;  
Best Local Similarity 100.0%; Pred. No. 8.5e-18;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISEADGKGANARGDKSIAIGDIAQALGSGSIAIGD 39  
DB 2 IGISEADGKGANARGDKSIAIGDIAQALGSGSIAIGD 40

## RESULT 4

US-08-968-685A-5  
Sequence 5, Application US/08968685A  
Patent No. 6214981

## GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSILA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-968-685A-5

Query Match 57.9%; Score 121; DB 4; Length 24;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28  
|||||  
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 5  
US-08-968-685A-13  
Sequence 13, Application US/08968685A  
Patent No. 6214981

GENERAL INFORMATION:  
APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSILA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-13

Query Match 57.9%; Score 121; DB 4; Length 24;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28  
|||||  
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 6  
US-08-968-685A-19  
Sequence 19, Application US/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSILA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-19

Query Match 57.9%; Score 121; DB 4; Length 24;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28  
|||||  
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 7  
US-09-268-347-49  
Sequence 49, Application US/09268347  
Patent No. 6335182  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
FILE REFERENCE: 1038-860  
CURRENT APPLICATION NUMBER: US/09/268,347  
CURRENT FILING DATE: 1999-03-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 49  
LENGTH: 2314  
TYPE: PRP  
ORGANISM: Haemophilus influenzae  
US-09-268-347-49

Query Match 43.1%; Score 90; DB 4; Length 2314;  
Best Local Similarity 52.3%; Pred. No. 0.0047;  
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGKGANAR---GDKSIAIGDIAIGSOSIAIGDKIV 43  
||||| |:::||||| ||| | ||||| :|

Db 2040 AKADGEAAVAIGRGTQAGNOSIAIGDMAQATGDSIAIGTGVNV 2083

RESULT 8  
US-09-268-347-48  
; Sequence 48, Application US/09268347  
; Patent No. 6335182  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347  
; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 2048  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-268-347-48

Query Match 42.6%; Score 89; DB 4; Length 2048;  
Best Local Similarity 52.3%; Pred. No. 0.0055;  
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGKANAR---GDKSIAIGDIAQALGSOSIAIGDKIV 43  
Db 1774 AKADGEAAVAIGRGTQAGNOSIAIGDMAQATGDSIAIGTGVNV 1817

RESULT 9  
US-08-188-582-9  
; Sequence 9, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; NUMBER OF SEQUENCES: 36  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 278 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-188-582-9

Query Match 28.7%; Score 60; DB 1; Length 278;  
Best Local Similarity 31.6%; Pred. No. 2.6;  
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

OY 1 IGISEADGKGKANARGDKSIAIGDIAQALGSOSIAIG 38  
Db 198 VGSSGGSGGGGGGQEVKSESTGAGDDLKMEVDSDAANG 235

RESULT 10  
US-08-646-715-9  
; Sequence 9, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; NUMBER OF SEQUENCES: 36  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582  
; FILING DATE: 28-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 278 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-646-715-9

Query Match 28.7%; Score 60; DB 1; Length 278;  
Best Local Similarity 31.6%; Pred. No. 2.6;  
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

OY 1 IGISEADGKGKANARGDKSIAIGDIAQALGSOSIAIG 38



Db 198 VGSSGGSGGGGDEYKSESTGAGGDLMEVDSPAANG 235

## RESULT 11

US-09-180-827-12

Sequence 12, Application US/09180827  
Patent No. 6355464  
GENERAL INFORMATION:  
APPLICANT: Healy, Judith M.  
APPLICANT: Bodorova, Jana  
APPLICANT: Lam, Kelvin T.  
APPLICANT: Lesson, Andrea J.  
TITLE OF INVENTION: M. Tuberculosis RNA Polymerase Alpha  
FILE REFERENCE: 0342/1C382-US2  
CURRENT APPLICATION NUMBER: US/09/180,827  
CURRENT FILING DATE: 1999-01-26  
PRIOR APPLICATION NUMBER: PCT/US97/22216  
PRIOR FILING DATE: 1997-11-03  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 379  
TYPE: PRT  
ORGANISM: C. trachomatis  
US-09-180-827-12

## Query Match

26.3% Score 55; DB 4; Length 379;  
Best Local Similarity 27.0%; Pred. No. 16;  
Matches 10; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 3 ISEADGKGANARGDKSIAIGDIAQALGSGSIAIGD 39

Db 106 LQDCGRCGRCOKLRATISVDASDLAAGGQKEVTIGD 142

## RESULT 12

US-09-336-447A-15

Sequence 15, Application US/09336447A  
Patent No. 6310190  
GENERAL INFORMATION:  
APPLICANT: HANSEN, ERIC J.  
APPLICANT: AEBL, CHRISTOPH  
APPLICANT: COPE, LESLIE D.  
APPLICANT: MACIVER, ISOBEL  
APPLICANT: FISKE, MICHAEL J.  
APPLICANT: FREDENBURG, ROSS A.  
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS  
FILE REFERENCE: AMCY:024  
CURRENT APPLICATION NUMBER: US/09/336,447A  
CURRENT FILING DATE: 1999-06-21  
NUMBER OF SEQ ID NOS: 98  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 15  
LENGTH: 889  
TYPE: PRT  
ORGANISM: Moraxella catarrhalis  
US-09-336-447A-15

## Query Match

26.3% Score 55; DB 4; Length 889;  
Best Local Similarity 34.1%; Pred. No. 43;  
Matches 14; Conservative 5; Mismatches 16; Indels 6; Gaps 1;

QY 9 GKGG-----ANARGDKSIAIGDIAQALGSGSIAIGD 43

Db 320 GKGSFAGIDNKRANADNAVALGNKNTIEGNSVALGSSNTIV 360

## RESULT 13

US-08-405-175A-6

Sequence 6, Application US/08405175A  
Patent No. 5885772

GENERAL INFORMATION:  
APPLICANT: Aderem, Alan A.  
APPLICANT: Chen, Jianmin  
APPLICANT: Chang, Sandy  
TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,175A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-121A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521

## INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 335 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: predicted primary structure of bovine MARCKS  
HYPOTHEICAL: NO  
US-08-405-175A-6

## Query Match

25.6% Score 53.5; DB 2; Length 335;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

QY 5 EADGKGANARGDKSIAIGDIAQALGSGSIAIGD 39

Db 182 EAEGA-AGASAGSGKDEASGGAAGAGAGAPGE 215

## RESULT 14

US-08-808-599A-24

Sequence 24, Application US/08808599A  
Patent No. 6111089  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-1A 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1160 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-599A-24

```

```

Query Match      25.1%; Score 52.5; DB 3; Length 1160;
Best Local Similarity 38.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 3; Mismatches 14; Indels 9; Gaps 2;

OY 2 GISPADGKGANARGDKSIAIGDI---AQLGSGSIAIGD 39
Db 171 GISNPGSGGGRN-----SIFGSPVNTSANSAPSISFGD 207

```

```

RESULT 15
US-08-173-508-6
Sequence 6, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136

```

```

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-508-6

```

```

Query Match      24.9%; Score 52; DB 1; Length 513;
Best Local Similarity 32.4%; Pred. No. 54;
Matches 11; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 GISPADGKGANARGDKSIAIGDIAQLGSGSI 35
Db 24 GSDPDNDGGRSAGPSAAPPSCVPALASQTL 57

```

Search completed: July 30, 2002, 15:34:51  
Job time: 236 sec

Wed Jul 31 07:54:03 2002

us-09-813-214a-1.ral



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:31:35 ; Search time 51.53 Seconds  
(without alignments)  
80.183 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGKGGKGNANRAGDS.....GDLAQLGSGSIAIGDNKIV 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.5	33.7	1588	2	probable adhesin Z
2	70.5	33.7	1588	2	probable adhesin E
3	66	31.6	1004	2	surface-exposed on
4	65.5	31.3	434	2	invasin precursor
5	65.5	31.3	658	2	probable surface P
6	64.5	30.9	1190	2	surface protein XF
7	64	30.6	1107	2	probable autotrans
8	63.5	30.4	455	2	Yopa protein - Yer
9	63	30.1	2059	2	surface protein XF
10	62.5	29.9	436	2	probable solute-bl
11	61	29.2	394	2	DNA-directed RNA P
12	61	29.2	615	2	hypothetical glyci
13	61	29.2	694	2	hypothetical glyci
14	60	28.7	278	2	transcription init
15	60	28.7	327	2	hypothetical prote
16	59.5	28.5	104	2	Chaperonins 10 kD
17	59.5	28.5	365	2	cell surface prote
18	59	28.2	401	2	protein C05B5.3 [I
19	58.5	28.0	562	2	related to merozoi
20	58	27.8	293	2	serine proteinase
21	58	27.8	1176	2	ice nucleation pro
22	57.5	27.5	1676	1	immediate-early pr
23	57.5	27.3	1660	2	hypothetical glyci
24	57	27.3	364	2	probable exported
25	57	27.3	422	2	Yopa protein - Yer
26	57	27.3	505	2	cell surface prote
27	56.5	27.0	1737	2	unconventional myo
28	56.5	27.0	245	2	conserved hypothet
29	56.5	27.0	2342	2	fatty-acid synthas

30	56	26.8	377	2	A71505	DNA-directed RNA P
31	56	26.8	552	2	T23755	hypothetical prote
32	56	26.8	584	2	G70804	hypothetical glyci
33	55.5	26.6	698	2	C82332	translational elonga
34	55	26.3	229	2	AB0220	flagellar L-ring P
35	55	26.3	447	2	E83465	conserved hypothet
36	54.5	26.1	297	2	E97387	hypothetical prote
37	54.5	26.1	340	2	AF2605	oxidoreductase Act
38	54.5	26.1	308	2	T20807	hypothetical prote
39	54.5	26.1	467	2	P96773	geranylgeranyl red
40	54.5	26.1	562	2	B70953	hypothetical glyci
41	54.5	26.1	644	2	T15652	hypothetical prote
42	54.5	26.1	728	2	AF3299	malate synthase 9
43	54.5	26.1	741	2	G70917	hypothetical glyci
44	54.5	26.1	1538	2	H70846	hypothetical glyci
45	54	25.8	196	2	G85435	TIN-1-like protein

## ALIGNMENTS

RESULT 1  
A:probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: A86036  
R:Perna, N.T., Plunkett III, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., May  
Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., DiMaio, E., Potamianos, K., Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A86036  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1588 <STO>  
A:Cross-references: GB:AE005174; NID:912518349; PIDN:AAG58749.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetic: 73;  
A:Gene: Z5029

Query Match 33.7%; Score 70.5; DB 2; Length 1588;  
Best Local Similarity 32.8%; Pred. No. 3.4;  
Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

QY 4 SPADG-----GKG-----GANNRGKSNIGDIAQLGSGSIAIGDNKI 42  
DB 261 NKADGVDAIAGNGSQSRGLNTIALGTASNATGDKSLAGSNSSANGINSVALGADSI 338

RESULT 2  
H91188  
probable adhesin ECS4480 [similarity] - Escherichia coli (strain O157:H7, substrain R  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 02-Nov-2001  
R:Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C  
gasawara, N., Yasunaga, T., Kuwara, S., Shiba, T., Hattori, M., Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: H91188  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1588 <NAV>  
A:Cross-references: GB:BA000007; PIDN:BA837903.1; PID:913363955; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECS4480

Query Match 33.7%; Score 70.5; DB 2; Length 1588;

Best Local Similarity 32.8%; Pred. No. 3.4;  
Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

QY 4 SEANG-----GKG-----GANARGDKSIADIDIAQALGSOSIADGNKI 42  
Db 281 NKADGVNIALGNSOSRGINTIALGTASNATGDKSLIALGSSNANGINSVALGADSI 338

RESULT 3  
C82672

surface-exposed outer membrane protein Xf1516 [imported] - *Xylella fastidiosa* (strain 9a)

C:Species: *Xylella fastidiosa*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: C82672

A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-1004 <SIM>  
A:Cross-references: GB:AE003961; GB:AE003849; NID:g9106543; PIDN:AAF84325.1; GSPDB:GN001

A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laiz

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, H

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saccelli, R.V.; Sawasak

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, R.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:

A:Gene: Xf1516

Query Match 31.6%; Score 66; DB 2; Length 1004;  
Best Local Similarity 50.0%; Pred. No. 7.3;

Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 9 GKGANARGDKSIADIDIAQALGSOSIATG 38  
Db 576 GSGVSAIKRSTATGASNAQAVGSSVALG 605

RESULT 4  
S04534

invasin precursor - *Yersinia pseudotuberculosis* plasmid pIBI

C:Species: *Yersinia pseudotuberculosis*  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999

C:Accession: S04534; S04910  
R:Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.

Nature 334, 522-525, 1988  
A:Title: Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations

A:Reference number: S04534; MUID:88302441  
A:Accession: S04534

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-434 <ROS>

A:Cross-references: EMBL:X12758; EMBL:X13883; NID:g48659; PIDN:CAA32088.1; PID:g48640

R:Skurnik, M.; Wolf-Watz, H.  
Mol. Microbiol. 3, 517-529, 1989  
A:Title: Analysis of the *yopA* gene encoding the *YopI* virulence determinants of *Yersinia*

A:Molecule type: DNA  
A:Residues: 1-434 <SKU>

A:Cross-references: EMBL:X12758; EMBL:X13883; NID:g48659; PIDN:CAA32088.1; PID:g48640

C:Genetics:  
A:Gene: *yopA*  
A:Genome: plasmid

Query Match 31.3%; Score 65.5; DB 2; Length 434;  
Best Local Similarity 46.5%; Pred. No. 3.6;

Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGANARGDK--SIADIDIAQ-----ALGSOSIADGNKI 42  
Db 94 GAGGILNARKDPRYSIATGTAAPAAVAVGSGSIATGVNSV 136

RESULT 5  
AH0110

probable surface protein (partial) YP00902 [imported] - *Yersinia pestis* (strain CO92)

C:Species: *Yersinia pestis*  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C:Accession: AH0110  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M

deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barril

Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0110

A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-658 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175

C:Genetics:  
A:Gene: YP00902

Query Match 31.3%; Score 65.5; DB 2; Length 658;  
Best Local Similarity 39.0%; Pred. No. 5.5;

Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 2 GISEADGGKGANARGDKSIADIDIAQALGSOSIADGNKI 42  
Db 480 GANSATGGAGSV-ASGNNSTAFSGAKATANSALANSV 519

RESULT 6  
A82615

surface protein Xf1981 [imported] - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: A82615  
R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82615

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1190 <SIM>

A:Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN

A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

Submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XFI981

Query Match 30.9%; Score 64.5; DB 2; Length 1190;  
Best Local Similarity 44.4%; Pred. No. 13;  
Matches 16; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

QY 4 SEADGGKGAN-ARGDKSIAIGDIAQALGSOSIAIG 38  
Db 191 SKTDGNTGATVAGQALRSIAIGTARSSQSDAISIG 226

RESULT 7  
AC0976  
probable autotransporter saps [imported] - Salmonella enterica subsp. enterica serovar  
C:Species: Salmonella enterica subsp. enterica serovar typh  
A:Note: this species has also been called Salmonella typh  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AC0976  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Mout, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AC0976  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1107 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD03303.1; PID:916504923; GSPDB:GN00176  
C:Genetics:  
A:Gene: saps

Query Match 30.6%; Score 64; DB 2; Length 1107;  
Best Local Similarity 38.5%; Pred. No. 14;  
Matches 15; Conservative 7; Mismatches 15; Indels 2; Gaps 1;

QY 4 SEADGGKGANARGDKSIAIGDIAQALGSOSIAIGDNKI 42  
Db 879 TMTDG--ADANAQADSVAGISGSIAMAAENVALCTNSV 915

RESULT 8  
S04912  
yopA protein - Yersinia enterocolitica plasmid pY6471/76  
C:Species: Yersinia enterocolitica  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S04912  
R:Skurnik, M.; Wolf-Watz, H.  
Mol. Microbiol. 3, 517-529, 1989  
A:Title: Analysis of the yopA gene encoding the Yop1 virulence determinants of Yersinia  
A:Reference number: S04910; MUID:89343638  
A:Accession: S04912  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <SKD>  
A:Cross-references: EMBL:X13882; NID:948606; PIDN:CAA32086.1; PID:948607  
C:Genetics:  
A:Gene: plasmid pY6471/76

Query Match 30.4%; Score 63.5; DB 2; Length 455;  
Best Local Similarity 44.2%; Pred. No. 6.4;  
Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

QY 9 GKGG--ANARGDKSIAIGDIAQALGSOSIAIGDNKI 42  
Db 59 GAGGLNMSAGHSISIAIGTATPAKAKAAVAGAGSIATGVNSV 101

RESULT 9  
D82671  
surface protein XFI529 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82671  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A8515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82671  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2059 <SIM>  
A:Cross-references: GB:AE003982; GB:AE003849; NID:9106554; PIDN:AAF4338.1; GSPDB:GN  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Relbach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; La  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miraca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XFI529

Query Match 30.18%; Score 63; DB 2; Length 2059;  
Best Local Similarity 36.7%; Pred. No. 34;  
Matches 18; Conservative 4; Mismatches 19; Indels 8; Gaps 1;

QY 2 GISEADGGK-----GANARGDKSIAIGDIAQALGSOSIAIGDNKI 42  
Db 1731 GNYGDGATGSKAIAAGVGTQASEGAAYVSGAAGSAGKSTAGRAAI 1779

RESULT 10  
T36478  
probable solute-binding protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36478  
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
submitted to the EMBL data library, September 1999  
A:Reference number: Z21607  
A:Accession: T36478  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-436 <SEED>  
A:Cross-references: EMBL:AL110470; PIDN:CAB54175.1; GSPDB:GN00070; SCOEDB:SCF85.18  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCF85.18

Query Match 29.9%; Score 62.5; DB 2; Length 436;  
Best Local Similarity 35.8%; Pred. No. 8.1;  
Matches 19; Conservative 5; Mismatches 16; Indels 13; Gaps 2;

QY 2 GISEADGG-----KGA-----NARGDKSIAIGDIAQALGSOSIAIGDNKI 41

```

Db      283 G1QADGCGFVSSGGIGYNNFPVPDGGKCDPSNAGNPNMQIYLSTISSKAIDEEK   335
                                                                    :|||::|||::|
RESULT  11
B81663
DNA-directed RNA polymerase, alpha chain TC0794 [Imported] - Chlamydia muridarum (strain B81663)
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C:Accession: B81663
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28; 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255
A:Accession: B81663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <TE>
A:Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39597.1; PID:g719082;
A:Experimental source: strain Nig9 (MOpn)
C:Genetics:
A:Gene: TC0794
C:Superfamily: DNA-directed RNA polymerase alpha chain

```

Query Match	Similarity	Score	DB	Length
Best Local	32.4%	Pred. No.	11	
Matches	12	Conservative	8	Mismatches 17; Indels 0; Gaps 0;

  

QY	3	ISEADGCKGGAAR	DRKSTAI	TDINQALG	SGSIAIGD	39
Db	123	LQPCBGRASQKRT	ATISDASD	LAAAGGQK	ITITVD	159

RESULT 12  
H70589  
hypothetical glycine-rich protein RV2853 - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 21-Jan-2000  
C:Accession: H70589  
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
N:Date: 5/3/98, 5/3/98, 5/3/98  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98259587  
A:Accession: H70589  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-615 <COL>  
A:Cross-references: GB:295207; GB:AL123456; NID:g3361745; PIDN:CAB08453.1; PID:c315182;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2853  
C:Superfamily: unassigned collagens

	29.2%	Score 61;	DB 2;	Length 615;
Query Match	41.5%;	Pred. No. 17;		
Best Local Similarity	4;	Mismatches 18;	Indels 2;	Gaps 1.
Matches	17;	Conservative		

  

QY	2	GISEADGGKGGANARSKDSTATG--DIAQALCSOSIAIGDN	40
		: : :	
Db	408	GGAGGNGGTGGACAPGAGCAGGAKKDINSLGDNATVYGCN	448

RESULT 13  
F70868  
hypothetical glycine-rich protein Rv2487c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: F70866:  
R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Randream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature,393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: AF0500; MUID:98295987  
A:Accession: F70866  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-694 <CO>  
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAAL6064.1; PID:g279  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2487c  
C:Superfamily: elastin

```
Query Match      29.2%; Score 61; DB 2; Length 694;  
Best Local Similarity 38.3%; Pcd No. 19;  
Matches 18; Conservative 3; Mismatches 16; Indels 10; Gaps 2
```

Oy      2    GISADGK-----CGANRDKSIALIGDIAQAL-----GSQSILAIg 38  
         |||||  
Db     GGAGGGGTAAAGGGGNGDGGAACGGTAAAFSGDDGNGSDSYAAG 338

RESULT 14  
A49067  
transcription initiation factor IID chain p42 - fruit fly (*Drosophila melanogaster*)  
N:Alternate names: TATA-binding protein (TBP)-associated factor TAFII40; TFIID complex  
C:Species: *Drosophila melanogaster*  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Nov-2000  
C:Accession: A49067; S42221  
R:Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A.; Tjian, R.  
Cell 75, 519-530, 1993  
A:Title: Drosophila TAF-II40 interacts with both a VPI6 activation domain and the basal  
A:Reference number: A49067; MUID:94037099  
A:Accession: A49067  
A:Status: preliminary  
A:Molecule type: mRNA  
N:Residues: 1-278 <GCO>  
A:Cross-references: GB:L29540; NID:g463048; PID:g463049  
A:Note: Parts of this sequence were confirmed by peptide sequencing  
R:Kokubo, T.; Gong, D.W.; Wootton, J.C.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.  
Nature 367, 484-487, 1994  
A:Title: Molecular cloning of *Drosophila* TFIID subunits.  
A:Reference number: S42220; MUID:94150630  
A:Accession: S42221  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-278 <KKC>  
A:Cross-references: EMBL:U06458; NID:g458679; PIDN:AAC47347.1; PID:g458680  
C:Genetics:  
A:Gene: FLYBase:Taf40  
A:Cross-references: FLYBase:FBgn0000617  
;Keywords: transcription initiation

Query_Match	28.7%	Score	60	DB	2	Length	278;
Best Local Similarity	31.6%	Pred.	No.	10,			
Matches	12; Conservative	8; Mismatches	18; Indels	0; Gaps	0;		
QY	1	TGISEADGGKGCANRGDKSTAGIDIAQLASQSQTATG	38				
	:	: : : : : : : : :	:	:	:	:	:
Db	198	VGSSSGSGGGGGGVNKSSESTGAGSLKKEVSDAAAV	235				

RESULT 15  
T00797  
hypothenical protein At2g32710 [imported] - Arabidopsis thaliana  
N;Alternate names: hypothenical protein F24L7.15



C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001  
C:Accession: T00797; E84736  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, February 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.  
A:Reference number: Z14204  
A:Accession: T00797  
A:Status: translated from GH/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-327 <R0U>  
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914702  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84736  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-327 <STO>  
A:Cross-references: GB:AE002093; NID:g2914702; PID:AMC04492.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g32710; F24L7.15  
A:Map position: 2  
A:Introns: 193/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 28.7%; Score 60; DB 2; Length 327;  
Best Local Similarity 33.3%; Pred. No. 12;  
Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;  
QY 2 GISEADGGCGNARGDKSIAGDIAQALGSGSTAI 37  
DB 12 GAGAGAGGGGGGGGESSIALMDVVPSSSSSLGV 47

Search completed: July 30, 2002, 15:35:50  
Job time: 255 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:32:46 ; Search time 27.4 Seconds  
(without alignments)  
60.764 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 1 IGISEADGKGGKGANANGDKS.....GDIAQLGSSQTAIGDNKIV 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65.5	31.3	434	1 YADA_YERPS	P10858 yersinia ps
2	63.5	30.4	455	1 YADA_YEREN	P31489 yersinia en
3	62.5	29.9	725	1 MAS2_PSEFL	O05137 pseudomonas
4	61	29.2	377	1 RPOA_CHLMT	O09194 chlamydia m
5	60	28.7	278	1 Y2D7_DROME	O27272 drosophila
6	59.5	28.5	104	1 CH10_SCHPO	O05904 schizosach
7	59	28.2	401	1 YKO3_CAEEL	P34291 caenorhabdi
8	57.5	27.5	676	1 ICP0_HSVBK	P29836 bovine herp
9	56.5	27.0	245	1 YH77_ARCFU	O28497 archaeoglob
10	56	26.8	377	1 RPOA_CHLTR	O46449 chlamydia t
11	55.5	26.6	498	1 NDD0_ALCXH	P94212 alcaligenes
12	54	25.8	1287	1 SK12_YEAST	P35207 saccharomyc
13	53.5	25.6	331	1 MACS_BOVIN	P12624 bos taurus
14	53.5	25.6	442	1 GLXD_RHIME	O87392 rhizobium m
15	53.5	25.6	699	1 EFG_HARIN	P43925 haemophilus
16	53.5	25.6	700	1 EFG_PASMU	P57838 pasteurella
17	53.5	25.6	703	1 EFG_ECOLI	P02996 escherichia
18	53.5	25.6	703	1 EFG_SALTY	P26229 salmonella
19	53.5	25.6	1043	1 DSG1_BOVIN	O03763 bos taurus
20	53	25.4	219	1 PGMB_ECOLI	P77366 escherichia
21	53	25.4	507	1 FLIC_SALBE	O06968 salmonella
22	53	25.4	507	1 FLIC_SALON	O06974 salmonella
23	53	25.4	595	1 PRIM_CHLTR	O84799 chlamydia t
24	52.5	25.1	107	1 YBDO_ECOLI	P75819 escherichia
25	52.5	25.1	329	1 HEM2_MYCLE	P46723 mycobacteri
26	52.5	25.1	702	1 EFG_BUCAL	P57593 buchnera ap
27	52	24.9	270	1 NUP1_PENCI	P24289 penicillium
28	52	24.9	270	1 NUP3_PENCO	P24504 penicillium
29	52	24.9	350	1 USRA_SALTY	P06196 salmonella
30	51.5	24.6	1901	1 EFG_THICU	O53553 mycobacteri
31	51.5	24.6	702	1 YZ08_MYCTU	O50565 thibacilli
32	51.5	24.6	2834	1 IP3R_DROME	P29993 drosophila
33	51	24.4	126	1 ANFC_MOUSE	O61839 mus musculu

34	51	24.4	434	1 COAA_BPKE	P03663 bacterioph
35	51	24.4	550	1 USHA_ECOLI	P07024 escherichia
36	51	24.4	580	1 IPAB_SHLDY	O03945 shigella dy
37	51	24.4	580	1 IPAB_SHFL	P18011 shigella fl
38	51	24.4	663	1 OTRA_STROM	O55002 streptomyce
39	50.5	24.2	136	1 ANFC_HUMAN	P23582 homo sapien
40	50.5	24.2	329	1 HEM2_MYCTU	O33357 mycobacteri
41	50.5	24.2	401	1 CSP_PLACG	P08674 plasmidium
42	50.5	24.2	488	1 FLIC_PSRFE	P72151 pseudomonas
43	50.5	24.2	513	1 ND11_YEAST	P32340 saccharomyc
44	50.5	24.2	655	1 UVRB_THERH	O56243 thermus aqu
45	50.5	24.2	676	1 ICP0_HSVBJ	P29128 bovine herp

## ALIGNMENTS

```

RESULT 1
ID YADA_YERPS STANDARD: PRT: 434 AA.
AC P10858;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Invasin precursor (Outer membrane adhesin).
CN YADA OR YOPA OR INVA OR YOPL.
OS Yersinia pseudotuberculosis.
OG Plasmid pIB1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YPIII;
RX MEDLINE=88302441; PubMed=3043229;
RA Rosqvist R., Skurnik M., Wolf-Watz H.;
RT "Increased virulence of Yersinia pseudotuberculosis by two
independent mutations."
RL Nature 334:522-525(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YPIII;
RX MEDLINE=89343638; PubMed=2761389;
RA Skurnik M., Wolf-Watz H.;
RT "Analysis of the yopa gene encoding the YopI virulence determinants
of Yersinia spp."
RL Mol. Microbiol. 3:517-529(1989).
DE FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
SURFACE.
DE -!- SUBCELLULAR LOCATION: Outer membrane.
DE This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL: X13883; CNA32088.1;
DR PIR: S04534; S04534.
DR PIR: S04910; S04910.
KW Plasmid; Virulence; Signal; Outer membrane.
FT SIGNAL 25
FT CHAIN 26 434 INVASIN.
SQ SEQUENCE 434 AA; 45054 MW; EE2C55FB12B183D4 CRC64;

```

Query Match 31.3%; Score 65.5; DB 1; Length 434;  
Best Local Similarity 46.5%; Pred. No. 1.9;

Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

Qy 9 GKGANARGDK--STAIGDIAO-----ALGSOSIAIDNKI 42  
 Db 94 GAGGLNRAKDPYSIALIGATTAETAKPAVAVGSGSITGVNSV 136

RESULT 2  
 YADA\_YEREN STANDARD; PRT; 455 AA.  
 AC P31489;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Invasin precursor (Outer membrane adhesin).  
 GN YADA OR YOPA OR INVA OR YOP1.  
 OS *Yersinia enterocolitica*.  
 CC Plasmid pY...  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Yersinia.  
 CC NCBI\_TaxID=630;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC SRRATN=6471/76 / SEROTYPE O:3;  
 CC MEDLINE=95020586; PubMed=7934875;  
 CC Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,  
 CC Skurnik M.;  
 CC "Hydrophobic domains affect the collagen-binding specificity and  
 CC surface polymerization as well as the virulence potential of the Yada  
 CC protein of *Yersinia enterocolitica*.";  
 CC Mol. Microbiol. 10:995-1011(1993).  
 CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO  
 CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE  
 CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS  
 CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL  
 CC SURFACE.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC CC  
 CC EMBL; X13882; CAA32086.1; -  
 CC DR PIR; S04912; S04912.  
 CC KW Plasmid; Virulence; Signal; Outer membrane.  
 CC FT SIGNAL 1 455 INVASIN.  
 CC FT CHAIN 26 455  
 CC FT SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;

Query Match 30.4%; Score 63.5; DB 1; Length 455;  
 Best Local Similarity 44.2%; Pred. No. 3.3;  
 Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

Qy 9 GKG--ANARGDKSIAIGDIAO-----ALGSOSIAIDNKI 42  
 Db 59 GAGGLNRAKDPYSIALIGATTAETAKPAVAVGSGSITGVNSV 101

RESULT 3  
 MASZ\_PSEFL STANDARD; PRT; 725 AA.  
 AC 005137;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Probable malate synthase G (EC 4.1.3.2).  
 GN GICB OR PC2.4.  
 OS *Pseudomonas fluorescens*.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas  
 CC NCBI\_TaxID=294;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=L6.5;  
 CC RA Artiguenave F.M., Delecu M., Vilagines R., Danglot C.;  
 CC "A functional glyoxylate bypass is mandatory for utilization of  
 CC alkanes by *Pseudomonas fluorescens*.";  
 CC Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC RL  
 CC -1- CATALYTIC ACTIVITY: L-malate + CoA = acetyl-CoA + H<sub>2</sub>O +  
 CC glyoxylate;  
 CC -1- PATHWAY: SECOND STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE  
 CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).  
 CC CC  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE G FAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC CC  
 CC EMBL; Y11998; CAA72726.1; -  
 CC DR HSSP; P37330; IDBC.  
 CC DR InterPro; IPR001465; Malate\_synthase.  
 CC KW Pfam; PF01274; Malate\_synthase; 1.  
 CC KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase.  
 CC SEQUENCE 725 AA; 78861 MW; 0C3326FE67C9B381 CRC64;

Query Match 29.9%; Score 62.5; DB 1; Length 725;  
 Best Local Similarity 40.7%; Pred. No. 6.8;  
 Matches 22; Conservative 2; Mismatches 17; Indels 13; Gaps 3;

Qy 3 ISEAG--GKGANARGDKSIAIG-----STAIGDIAO-----SOSIAIDNKI 43  
 Db 18 ISEAGAKRGKYNKVRGDKYIAFAFLDEAPLASGSHVDSYGYRTADCKLI 201

RESULT 4  
 RPOA\_CHLMU STANDARD; PRT; 377 AA.  
 AC QPBN4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (transcriptase  
 DE alpha chain) (RNA polymerase alpha subunit).  
 GN RPOA OR TC0794.  
 OS *Chlamydia muridarum*.  
 CC Chlamydia; Chlamydiales; Chlamydiaceae; Chlamydia.  
 CC NCBI\_TaxID=83560;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=MOPN / NIGS;  
 CC MEDLINE=20150235; PubMed=10684935;  
 CC Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 CC White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,  
 CC Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 CC Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 CC Eisen J., Fraser C.M.;  
 CC "Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia*  
 CC pneumoniae AR39.";  
 CC Nucleic Acids Res. 28:1397-1406(2000).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).



Db 198 VSSSGSGGGGGEVSESGAGDLMKMDSDAAVAG 235

# RESULT 6

CH10\_SCHPO STANDARD; PRT; 104 AA.

AC 059804;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 10 kDa heat shock protein, mitochondrial (HSP10) (10 kDa chaperonin).  
 GN HSP10 OR SPCC550.06C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;

RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;

CC Submitted (Mar-1998) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: EUKARYOTIC CPN10 HOMOLOG WHICH IS ESSENTIAL FOR  
 CC MITOCHONDRIAL PROTEIN BIOGENESIS, TOGETHER WITH CPN60, BINDS TO  
 CC CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES THE ATPASE ACTIVITY  
 CC OF THE LATTER (BY SIMILARITY).  
 CC -I- SUBUNIT: HOMOHXAMER (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -I- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AL023592; CAA19110.1; -  
 CC InterPro: IPR001476; CPN10.  
 CC Pfam: PF00166; cpn10.1.  
 CC DR PRINTS: PR00297; CHAPERONIN10.  
 CC DR PRODOM: PD000566; Cpn10; 1.  
 CC DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1.  
 CC DR Chaperone; Mitochondrion.  
 SQ SEQUENCE 104 AA; 11282 MW; 16BF6870BCB10222 CRC64;

Query Match 28.5%; Score 59.5; DB 1; Length 104;  
 Best Local Similarity 36.8%; Pred. No. 2.2;  
 Matches 14; Conservative 7; Mismatches 12; Indels 5; Gaps 2;

OY 9 GKGGANRGC---DKSIAGD--IAQALGSGSIAIGDNK 41  
 DB 52 GKGYNKGGKLAQPSVAVGDVLLPAYGGSNIKWGEER 89

# RESULT 7

YK03\_CAEEL STANDARD; PRT; 401 AA.

AC P34291;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypothetical 43.5 kDa protein C05B5.3 in chromosome III.  
 GN C05B5.3  
 OS Caenorhabditis elegans.

CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
 RA Mortimore B.J.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.

RA Durbin R.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z32679; CAA83596.1; -  
 CC DR WormPep; C05B5.3; CE17369.  
 KW Hypothetical protein.  
 SQ SEQUENCE 401 AA; 43472 MW; C99DF259EF6C8B55 CRC64;

Query Match 28.2%; Score 59; DB 1; Length 401;  
 Best Local Similarity 42.4%; Pred. No. 9.5;  
 Matches 14; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

OY 8 GKGGANRGC---DKSIAGD--IAQALGSGSIAIGDN 40  
 DB 339 GGGGSGGAPGSGSVPMTEAQAIVQIAIRDN 371

# RESULT 8

ICP0\_HSVBK STANDARD; PRT; 676 AA.

AC P29836;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trans-acting transcriptional protein ICP0 (P135 protein) (IER  
 DE 2.9/EK2.6).  
 GN BICP0.  
 OS Bovine herpesvirus type 1 (strain K22).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.

OX NCBI\_TaxID=31519;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=92219360; PubMed=1313901;  
 RA Wirth U.V., Fraefel C., Vogt B., Vicek C., Paces V., Schwyzer M.;  
 RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1  
 RT are 3' coterminal and encode a putative zinc finger transactivator  
 RT protein.";

RL J. VIROL. 66:2763-2772(1992).  
 CC -I- PFM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL  
 CC ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY  
 CC CASEIN KINASE II.

CC -I- SIMILARITY: CONTAINS 1 ZINC-FINGER ZINC FINGER.  
 CC -I- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M8464; AAA46061.1; -  
 CC DR PIR: A38209; EDBE22.  
 DR HSSP: P28980; ICHC.  
 DR InterPro: IPR001841; Znf\_fing.  
 DR Pfam: PF00097; zf-C3HC4; 1.  
 DR SMART; SM00184; RING; 1.

DR PROSITE: PS00518; 2F\_RING.1; 1.  
DR PROSITE: PS50089; 2F\_RING.2; 1.  
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;  
RW DNA-binding; Early protein; Repressor; Phosphorylation.  
FT ZN\_FING 13 52 RING-TYPE.  
FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).  
SO SEQUENCE 676 AA; 67701 MW; 99B0683C9BFC65D CRC64;

Query Match 27.5%; Score 57.5; DB 1; Length 676;  
Best Local Similarity 44.1%; Pred. No. 24;  
Matches 15; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

Db 125 EYKVGKIGFSAKLEKATVGEIAEKLGPAWLPFGERV 163

RESULT 10  
ID RPOA\_CHLTR STANDARD; PRT; 377 AA.  
AC 046449; 084515;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase  
DE alpha-chain) (RNA polymerase alpha subunit).  
GN RPOA OR CT507.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L2/434/BU;  
RX MEDLINE=95247702; PubMed=773029;  
RA Gu L.J., Wenman W.M., Remacha M., Meuser R., Coffin J., Kaul R.;  
RT "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and  
RT structural analysis."  
RL J. Bacteriol. 177:2594-2601(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/W-3/CC;  
RX MEDLINE=99008093; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis."  
RL Science 282:754-759(1998).  
RC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA)(N).  
CC -1- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH  
IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA CHAIN AND 1  
OMEGA CHAIN (BY SIMILARITY).  
CC -1- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF  
CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH  
TRANSCRIPTIONAL REGULATORS.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L33834; AAA74989.1; .  
DR EMBL: AE001323; AAC68108.1; .  
DR HSSP: P00574; ICOO.  
DR Sienna-2DPAGE: Q46449; .  
DR InterPro: IPR001700; RNA-pol\_A\_bac.  
DR Pfam: PF01000; RNA-pol\_A\_bac; 1.  
DR Pfam: PF03118; RNA-pol\_A\_CTD; 1.  
DR Pfam: PF001179; RNA-pol\_A\_bac; 1.  
KW Transcriptase; Transcription; DNA-directed RNA polymerase;  
KW Complete proteome.  
FT VARIANT 30 G -> S (IN SEROVAR L2).  
FT VARIANT 92 IV -> ML (IN SEROVAR L2).  
FT VARIANT 124 I -> V (IN SEROVAR L2).  
FT VARIANT 377 G -> A (IN SEROVAR L2).  
FT VARIANT 377 G -> A (IN SEROVAR L2).  
FT VARIANT 295 R -> R (IN REF. 2).  
FT VARIANT 295 R -> R (IN REF. 2).  
SO SEQUENCE 377 AA; 41824 MW; 09B55FBBC36804A CRC64;

DR PROSITE: PS00518; 2F\_RING.1; 1.  
DR PROSITE: PS50089; 2F\_RING.2; 1.  
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;  
RW DNA-binding; Early protein; Repressor; Phosphorylation.  
FT ZN\_FING 13 52 RING-TYPE.  
FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).  
SO SEQUENCE 676 AA; 67701 MW; 99B0683C9BFC65D CRC64;

Query Match 27.0%; Score 56.5; DB 1; Length 245;  
Best Local Similarity 33.3%; Pred. No. 11;  
Matches 13; Conservative 12; Mismatches 13; Indels 1; Gaps 1;

Db 5 EADGKGANARQDKSIAIGDIAQALG-SQSIAIGDKKI 42

RESULT 9  
ID YH77\_ARCFU STANDARD; PRT; 245 AA.  
AC 028497;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF1777.  
GN AF1777.  
OS Archaeoglobus fulgidus.  
OC Archaeae; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OC Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Peterson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirtress E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodde A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton P.W., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus."  
RL Nature 390:364-370(1997).  
CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE000980; AAB89474.1; .  
DR TIGR: AF1777; .  
DR InterPro: IPR002678; DUF34.  
DR Pfam: PF01784; DUF34; 1.  
KW Hypothetical protein; Complete proteome.  
SO SEQUENCE 245 AA; 26860 MW; AA516IDCF10B04D CRC64;

Query Match 26.8%; Score 56; DB 1; Length 377;  
 Best Local Similarity 29.7%; Pred. No. 20;  
 Matches 11; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

3 ISEADGKGGKGANMAGKMSIALGDAOLGSOAICD 39  
 106 LQDEGGRCSOKLRTATISIDASDLAAGGKEVTLGD 142

RESULT 11  
 NDDO\_ALCXX STANDARD; PRT; 498 AA.

AC P94212;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE N-acyl-D-aspartate deacylase (EC 3.5.1.83) (N-acyl-D-aspartate  
 deamidohydrolase).  
 OS Alkaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).  
 OC Bacteria; Proteobacteria; beta subdivision; Alkaligenaceae;  
 OX Achromobacter.  
 RN NCBI\_TaxID=515;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A-6;  
 RA Wakayama M., Watanabe E., Takenaka Y., Miyamoto Y., Tau Y., Sakai K.,  
 RA Moriyuchi M.;  
 RT "Cloning, expression, and nucleotide sequence of the N-acyl-D-  
 aspartate amidohydrolase gene from Alkaligenes xylosoxydans subsp.  
 RT xylosoxydans A-6.";  
 RL J. Ferment. Bioeng. 80:311-317(1995).  
 RN [2]  
 RP CHARACTERIZATION.  
 RC STRAIN-A-6;  
 RX MEDLINE=93372486; PubMed=7763985;  
 RA Moriyuchi M., Sakai K., Katsumo Y., Maki T., Wakayama M.;  
 RT Purification and characterization of novel N-acyl-D-aspartate  
 RT amidohydrolase from Alkaligenes xylosoxydans subsp. xylosoxydans  
 RT A-6.";  
 RL Biosci. Biotechnol. Biochem. 57:1145-1148(1993).  
 CC -1- CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O = carboxylate + D-  
 CC aspartate.  
 CC -1- COFACTOR: ZINC.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: D45919; BAA08350.1; -  
 KW Hydrolase; Zinc.  
 SQ SEQUENCE 498 AA; 53584 MW; EAF6662891F9F1B CRC64;

Query Match 26.8%; Score 55.5; DB 1; Length 498;  
 Best Local Similarity 40.0%; Pred. No. 29;  
 Matches 14; Conservative 4; Mismatches 12; Indels 5; Gaps 1;

2 GISEADGKGGKGAN-----ARGDKSTAIGDAOLG 31  
 20 GATLIDGGGGPARGGDLAVRGGRIVYALGDFAPAPG 54

RESULT 12  
 SK12\_YEAST STANDARD; PRT; 1287 AA.  
 AC P35207; 006047;  
 DT 01-FEB-1994 (Rel. 28, Created)

30-MAY-2000 (Rel. 39, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Antiviral protein SK12.  
 GN SK12 OR YLR398C OR L8084.17.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93309467; PubMed=8321235;  
 RA Widner W.R., Winkler R.B.;  
 RT "Evidence that the SKI antiviral system of Saccharomyces cerevisiae  
 RT acts by blocking expression of viral mRNA.";  
 RL Mol. Cell. Biol. 13:4331-4341(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RX MEDLINE=97313267; PubMed=9169871;  
 RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansgore W.,  
 RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,  
 RA Eutian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,  
 RA Heuss-Weltzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,  
 RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,  
 RA Muller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,  
 RA Portetelle D., Purnelle B., Rechemann S., Rieger M., Rinke M., Rose M.,  
 RA Scharfe M., Scherens B., Scholler P., Schwarz C., Schwarz S.,  
 RA Underwood A.P., Uristarazu L.A., Vandenbol M., Verhasselt P.,  
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohnselt J.D.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 RT Nature 387:87-90(1997).  
 RN [3]  
 RP SEQUENCE OF 1-162 FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE=95116323; PubMed=7816623;  
 RA Lygeron Z., Conesa C., Lesage P., Swanson R.N., Ruat A., Carlson M.,  
 RA Sentenac A., Seraphin B.;  
 RT "The yeast Bpfl gene encodes a transcription factor involved in the  
 RT expression of a broad class of genes including snRNAs.";  
 RL Nucleic Acids Res. 22:5332-5340(1994).  
 CC -1- FUNCTION: REPRESENTS DSRNA VIRUS PROPAGATION BY SPECIFICALLY  
 CC BLOCKING TRANSLATION OF VIRAL MRNAS. PERHAPS RECOGNIZING THE  
 CC ABSENCE OF CAP OR POLY(A). ESSENTIAL FOR CELL GROWTH ONLY IN THE  
 CC PRESENCE OF M1 REPLICON. SEEMS TO BE A HELICASE.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: L13469; AAA35049.1; -  
 DR EMBL: U19729; AAB82356.1; -  
 DR EMBL: Z18944; CAA79378.1; -  
 DR PIR: B48140; B48140.  
 DR SGD: S0004390; SK12.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF002270; DEAD\_1.  
 DR Pfam: PF00271; helicase\_C\_1.  
 DR SMART: SM00487; DEXDC\_1.  
 DR SMART: SM00490; HELIC\_C\_1.  
 KW Antiviral; Translation regulation; RNA-binding; DNA-binding; Helicase;  
 KW ATP-binding.  
 FT NP\_BIND 351 358 ATP (POTENTIAL).  
 FT SITE 444 447 DEWV BOX.  
 FT DOMAIN 556 577 RNA-BINDING RGG-BOX (BY SIMILARITY).  
 FT DOMAIN 555 597 ARG/GLY-RICH.



FT CONFLICT 326 326 W -> C (IN REF. 1).  
 FT CONFLICT 759 760 OM -> L (IN REF. 1).  
 SQ SEQUENCE 1287 AA; 146058 MW; 7CCD36CFCD0F6C32 CRC64;

Query Match  
 Best Local Similarity 40.9%; Pred. No. 1.1e+02;  
 Matches 18; Conservative 5; Mismatches 17; Indels 4; Gaps 2;

OY 2 GISEADGKGAN--ARCDKSIAGDIAQALGSGST-ATGDNK 41  
 DB 557 GGTARGGRGSGNTRDGRGGRNSTRGANGSGRGAIGSNK 600

RESULT 13  
 MACS\_BOVIN STANDARD: PRT; 331 AA.  
 AC P12624;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Myristoylated alanine-rich C-kinase substrate (MARCKS) (ACAMP-81).  
 GN MACS.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89282412; PubMed=273411;  
 RA Stumpo D.J., Graff J.M., Albert K.A., Greengard P., Blackshear P.J.;  
 RT "Nucleotide sequence of a cDNA for the bovine myristoylated  
 RT alanine-rich C kinase substrate (MARCKS).";  
 RL Nucleic Acids Res. 17:3987-3988(1989).  
 RN (2)  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=89264553; PubMed=2726763;  
 RA Stumpo D.J., Graff J.M., Albert K.A., Greengard P., Blackshear P.J.;  
 RT "Molecular cloning, characterization, and expression of a cDNA  
 RT encoding the '80- to 87-kDa' myristoylated alanine-rich C kinase  
 RT substrate: a major cellular substrate for protein kinase C.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4012-4016(1989).  
 RN (3)  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=92171958; PubMed=1540183;  
 RA Mizutani A., Tokumitsu H., Hidaka H.;  
 RT "Acidic calmodulin binding protein, ACAMP-81, is MARCKS protein  
 RT interacting with synapsin I.";  
 RL Biochem. Biophys. Res. Commun. 182:1395-1401(1992).  
 RN (4)  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE=89308594; PubMed=2473066;  
 RA Graff J.M., Stumpo D.J., Blackshear P.J.;  
 RT "Characterization of the phosphorylation sites in the chicken and  
 RT bovine myristoylated alanine-rich C kinase substrate protein, a  
 RT prominent cellular substrate for protein kinase C.";  
 RL J. Biol. Chem. 264:11912-11919(1989).  
 RN (5)  
 RP PHOSPHORYLATION SITES, AND REVISIONS.  
 RC TISSUE=Brain;  
 RX MEDLINE=94308052; PubMed=8034575;  
 RA Taniguchi H., Manenti S., Suzuki M., Titani K.;  
 RT "Myristoylated alanine-rich C kinase substrate (MARCKS), a major  
 RT proline-directed protein kinase(s), is an in vivo substrate of  
 RT the post-translational modifications.";  
 RL J. Biol. Chem. 269:18299-18302(1994).  
 RN (6)  
 RP REVERSIBLE ASSOCIATION WITH THE MEMBRANE.  
 RX MEDLINE=9138951; PubMed=2034276;  
 RA Thelen M., Rosen A., Nairn A.C., Aderem A.;  
 RT "Regulation by phosphorylation of reversible association of a

RT myristoylated protein kinase C substrate with the plasma membrane.";  
 RL Nature 351:320-322(1991).  
 RN (7)  
 RP ACTIN-FILAMENT CROSS-LINKING.  
 RX MEDLINE=92220195; PubMed=1560845;  
 RA Hartwig J.H., Thelen M., Rosen A., Janney P.A., Nairn A.C.,  
 RA Aderem A.;  
 RT "MARCKS is an actin filament crosslinking protein regulated by  
 RT protein kinase C and calcium-calmodulin.";  
 RL Nature 356:618-622(1992).  
 CC -1- FUNCTION: MARCKS IS THE MOST PROMINENT CELLULAR SUBSTRATE FOR  
 CC PROTEIN KINASE C. THIS PROTEIN BINDS CALMODULIN, ACTIN, AND  
 CC SYNAPSIN. MARCKS IS A FILAMENTOUS (F) ACTIN CROSS-LINKING PROTEIN.  
 CC -1- PPM: PHOSPHORYLATION BY PKC REPLACES MARCKS FROM THE MEMBRANE. IT  
 CC ALSO INHIBITS THE F-ACTIN CROSS-LINKING ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE MARCKS FAMILY.  
 CC -1- CAUTION: REF.1 AND REF.2 SEQUENCES DIFFER FROM THAT SHOWN IN  
 CC POSITIONS 140 TO 150 DUE TO A FRAMESHIFT.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, M24638; AAA30635.1; ALT\_FRAME.  
 CC PIR, A52904; A52904.  
 CC PIR, P50338; P50338.  
 CC PIR, S08341; S08341.  
 CC InterPro: IPR002101; MARCKS.  
 CC Pfam: PF02063; MARCKS.1.  
 CC PRINTS: PR00963; MARCKS.  
 CC PROSITE: PS00826; MARCKS\_1; 1.  
 CC PROSITE: PS00827; MARCKS\_2; 1.  
 CC KW Phosphorylation; Myristate; Calmodulin-binding; Actin-binding;  
 CC Membrane.  
 CC KW INT\_MET 0 0  
 CC FT LIPID 1 1 MYRISTATE.  
 CC FT DOMAIN 150 174 CALMODULIN-BINDING (PSD).  
 CC FT MOD\_RES 26 26 PHOSPHORYLATION.  
 CC FT MOD\_RES 45 45 PHOSPHORYLATION.  
 CC FT MOD\_RES 80 80 PHOSPHORYLATION.  
 CC FT MOD\_RES 99 99 PHOSPHORYLATION.  
 CC FT MOD\_RES 116 116 PHOSPHORYLATION.  
 CC FT MOD\_RES 133 133 PHOSPHORYLATION (BY PKC).  
 CC FT MOD\_RES 157 157 PHOSPHORYLATION (BY PKC).  
 CC FT MOD\_RES 161 161 PHOSPHORYLATION (BY PKC).  
 CC FT MOD\_RES 165 165 PHOSPHORYLATION (BY PKC).  
 CC FT MOD\_RES 168 168 PHOSPHORYLATION (BY PKC).  
 CC SQ SEQUENCE 331 AA; 31450 MW; 0F77BF117EDB35CA CRC64;

Query Match  
 Best Local Similarity 25.6%; Score 53.5; DB 1; Length 331;  
 Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

OY 5 EADGKGANANGDKSIAGDIAQALGSGSTAGD 39  
 DB 181 EAEGA-AGASAGKDEAGGAAAGGAAPGE 214

RESULT 14  
 GLXD\_RHIME STANDARD: PRT; 442 AA.  
 AC 087392;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Glutamate synthase large subunit-like protein.  
 GN GLXD OR R00089 OR SMC02612.  
 GN Rhizobium meliloti (Sinorhizobium meliloti).

CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Sinorhizobium.  
 CC NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RA Powers E.L., Vuyyuru V., Kahn M.L.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RA MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Driano S., Gloux S.,  
 RA Goffie T., Goffeau A., Kahn D., Kiss E., Leleire V., Maury D.,  
 RA Pohl T., Portetalle D., Puenhler A., Puenhler B., Ransperger U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallbert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 CC -1- SIMILARITY: TO GLUTAMATE SYNTHASES LARGE SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AF055582; AAC62222.1; -;  
 CC DR EMBL; AL591782; CAC41476.1; -;  
 CC DR InterPro: IPR003009; FMN\_enzyme.  
 CC DR InterPro: IPR002932; Glu\_synthase.  
 CC Pfam: PF01645; Glu\_synthase; 1.  
 CC Oxidoreductase; Complete proteome.  
 CC FT CONFLICT 322 322 A -> V (IN REF. 1).  
 CC FT CONFLICT 322 322 A -> V (IN REF. 1).  
 CC SQ SEQUENCE 442 AA; 47376 MW; CCF08B8EC2D61ACE CRC64;

Query Match 25.6%; Score 53.5; DB 1; Length 442;  
 Best Local Similarity 42.4%; Pred. No. 44;  
 Matches 14; Conservative 8; Mismatches 8; Indels 3; Gaps 2;

QY 11 GGANARGD--KSIATGDIQAQALGSGS--IAIGDN 40  
 DB 302 GGIRSGADVAKALAGADVAIGTAALVATGDN 334

RESULT 15  
 EFG\_HAEIN  
 ID EFG\_HAEIN STANDARD; PRT; 699 AA.  
 AC PA3925;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Elongation factor G (EF-G).  
 GN FUSA OR FUS OR HI0579.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 NC NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spillig T., Hedblom E., Cotton M.D.,  
 RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,

RA Gresham C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION  
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 CC RIBOSOME.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC -1- EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U32739; AAC22237.1; -;  
 CC DR HSSP; P13551; IFNM.  
 CC DR TIGR; HI0579; -;  
 CC DR InterPro: IPR000640; EFG\_C.  
 CC DR InterPro: IPR000795; GTP\_EFTU.  
 CC DR InterPro: IPR004161; GTP\_EFTU\_D2.  
 CC DR Pfam: PF06679; EFG\_C; 1.  
 CC DR Pfam: PF00009; GTP\_EFTU; 1.  
 CC DR Pfam: PF03144; GTP\_EFTU\_D2; 1.  
 CC DR PRINTS: PR00315; ELONGATNFCF.  
 CC DR PROSITE: PS00301; EFACITOR\_GTP; 1.  
 CC DR Elongation factor; Protein biosynthesis; GTP-binding;  
 CC Complete proteome.  
 CC KW INIT\_MER 0  
 CC FT NP\_BIND 16 23 GTP (BY SIMILARITY).  
 CC FT NP\_BIND 87 91 GTP (BY SIMILARITY).  
 CC FT NP\_BIND 141 144 GTP (BY SIMILARITY).  
 CC SQ SEQUENCE 699 AA; 77132 MW; FBBAD639C0F62801 CRC64;

Query Match 25.6%; Score 53.5; DB 1; Length 699;  
 Best Local Similarity 42.9%; Pred. No. 69;  
 Matches 12; Conservative 5; Mismatches 10; Indels 1; Gaps 1;

QY 13 ANARGD-KSIATGDIQAQALGSGSIAIGD 39  
 DB 365 ANKREIKERVAGDIAAALGILKDVTTGD 392

Search completed: July 30, 2002, 15:37:58  
 Job time: 312 sec

Wed Jul 31 07:54:05 2002

us-09-813-214a-1.rsp



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:31:55 ; Search time 88.56 Seconds  
(without alignments)  
83.997 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209  
Sequence: 1 IGISEADGKGGKGNARGDS.....GDIQAQLSGSIAIGNKIV 43

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	36.8	1299	16	Q9F3X6
2	72.5	34.7	1328	2	Q9LAX0
3	70	33.5	1953	16	Q98HJ2
4	68	32.5	2712	16	Q9F3X5
5	67	32.1	1265	2	Q9FDAO
6	67	32.1	1281	16	Q9ZK07
7	66	31.6	1004	16	Q9PD63
8	64.5	30.9	1190	16	Q9PC04
9	64	30.6	1107	2	Q9FZD8
10	63.5	30.4	454	2	Q85267
11	63	30.1	261	11	Q9EOT3
12	63	30.1	487	2	Q9LA53
13	63	30.1	2059	16	Q9PD50
14	62.5	29.9	436	2	Q9S2C3
15	61	29.2	550	2	Q9AVH8
16	61	29.2	567	2	Q9AFP8

17	61	29.2	615	16	005806	005806 mycobacteri
18	61	29.2	694	16	053212	053212 mycobacteri
19	60.5	28.9	698	5	Q9VMM5	Q9VMM5 drosophila
20	60	28.7	253	5	Q95R14	Q95R14 drosophila
21	60	28.7	286	10	Q9ACM0	Q9ACM0 arabidopsis
22	60	28.7	289	10	Q9FR83	Q9FR83 arabidopsis
23	60	28.7	327	10	Q48846	Q48846 arabidopsis
24	60	28.7	449	5	Q9V7C0	Q9V7C0 glycolysis
25	59.5	28.5	452	10	Q9XEP4	Q9XEP4 glycine max
26	59	28.2	511	9	Q9MCT8	Q9MCT8 bacterioph
27	58.5	28.0	388	3	Q9P557	Q9P557 neospora
28	58.5	28.0	464	10	Q9ZS34	Q9ZS34 nicotiana t
29	58	27.8	504	2	Q9LAS6	Q9LAS6 escherichia
30	58	27.8	1176	12	Q98587	Q98587 paramycium
31	57.5	27.5	403	4	Q9UHR6	Q9UHR6 homo sapien
32	57.5	27.5	1660	16	053215	053215 mycobacteri
33	57	27.3	422	2	056930	056930 yersinia en
34	57	27.3	422	2	Q93KR4	Q93KR4 yersinia en
35	57	27.3	737	2	Q9AF04	Q9AF04 frankia sp.
36	57	27.3	1737	5	Q9TW28	Q9TW28 dictyostell
37	57	27.3	2301	5	Q95ZD0	Q95ZD0 leishmania
38	56.5	27.0	2342	5	001677	001677 bombyx mori
39	56	26.8	369	2	Q9AF00	Q9AF00 bacillus sp
40	56	26.8	552	5	021538	021538 caenorhabdi
41	56	26.8	584	16	053575	053575 mycobacteri
42	55.5	26.6	698	16	Q9KU27	Q9KU27 vibrio chol
43	55	26.3	265	9	Q9M8T0	Q9M8T0 staphylococ
44	55	26.3	335	5	061538	061538 drosophila
45	55	26.3	399	16	Q98WS4	Q98WS4 rhizobium 1

## ALIGNMENTS

RESULT 1  
ID Q9F3X6 PRELIMINARY; PRT; 1299 AA.  
AC Q9F3X6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MAFK PROTEIN (HSF).  
GN MAFK OR HSF\_2 OR PM1570.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_taxid=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
RT "Evolutionary origins of the autotransporter proteins";  
RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145856; PubMed=11248100;  
RA May B.T., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Pm70";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
DR EMBL; AJ277635; CAC14202.1;  
KW Complete proteome.  
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 36.8%; Score 77; DB 16; Length 1299;  
Best Local Similarity 45.7%; Pred. No. 1.3;  
Matches 21; Conservative 7; Mismatches 12; Indels 6; Gaps 2;

QY 1 IGISEADGKGGK---GANKRDK--SIAIGDIAQLSGSIAIGNKIV 40  
DB 105 IGFGATNDGTNVAIGAKSKSKASISIAIGNAKALDQALTAIGN 154

RESULT 2  
 09LAXO  
 ID 09LAXO PRELIMINARY; PRT: 1328 AA.  
 AC 09LAXO;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOHERETICAL 127.4 KDA PROTEIN (FRAGMENT).  
 OS Xanthomonas campestris pv. pelargonii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 NCBI\_TaxID=91612;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=XPEL-1;  
 RA Kim J.F., Zumboff C.H., Beer S.V.;  
 RT "An alanine-, glycine-, and serine-rich protein and a putative serine  
 RT protease of Xanthomonas campestris pv. pelargonii."  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF083618; AAF63394.1; -  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 1328 AA; 127402 MW; ACE01A4C6548364C CRC64;

Query Match 34.7%; Score 72.5; DB 2; Length 1328;  
 Best Local Similarity 43.5%; Pred. No. 4.7;  
 Matches 20; Conservative 6; Mismatches 15; Indels 5; Gaps 2;  
 QY 2 GISEADG--GKGG--ANAGDKSIAIGDIAQALGSGSIAIGDKNI 42  
 Db 236 GLSTASGFLSSAGGYLSRASGASTAFGYARSESSSIAGVGTAL 281

RESULT 3  
 098HJ2  
 ID 098HJ2 PRELIMINARY; PRT: 1953 AA.  
 AC 098HJ2;  
 DT 01-OCT-2001 (TREMblrel. 18, Created)  
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
 DE MLL2848 PROTEIN.  
 GN MLL2848.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpou S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003000; BAB49874.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;

Query Match 33.5%; Score 70; DB 16; Length 1953;  
 Best Local Similarity 35.4%; Pred. No. 14;  
 Matches 17; Conservative 9; Mismatches 16; Indels 6; Gaps 1;  
 QY 1 IGISEADGKGG-----ANAGDKSIAIGDIAQALGSGSIAIGDKNI 42  
 Db 359 IGLGATAGGVDTAIGKQANASQAQDRIAMGTSIAKASSQAQIAVGINAV 406

RESULT 4  
 09F3X5  
 ID 09F3X5 PRELIMINARY; PRT: 2712 AA.  
 AC 09F3X5;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MAPB PROTEIN (HSE).  
 GN MAPB OR HSE\_1 OR PM0714.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
 RT "Evolutionary origins of the autotransporter proteins."  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida pm70."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL: AJ277636; CAC14203.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;

Query Match 32.5%; Score 68; DB 16; Length 2712;  
 Best Local Similarity 57.7%; Pred. No. 35;  
 Matches 15; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 17 GPKSTAIGDIAQALGSGSIAIGDKNI 42  
 Db 324 GESSVAIGDKAVSRGEASIAIGKNAI 349  
 RESULT 5  
 09FDA0  
 ID 09FDA0 PRELIMINARY; PRT: 1265 AA.  
 AC 09FDA0;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE PUTATIVE OUTER MEMBRANE PROTEIN XADA.  
 GN XADA.  
 OS Xanthomonas oryzae pv. oryzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 NCBI\_TaxID=64187;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ray S.K., Rajeshwari R., Sonti R.V.;  
 RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae  
 RT that is involved in virulence."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF288222; AAC01335.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;

Query Match 32.1%; Score 67; DB 2; Length 1265;  
 Best Local Similarity 46.2%; Pred. No. 20;  
 Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
 QY 13 ANAGDKSIAIGDIAQALGSGSIAIG 38  
 Db 694 AATGVSAAVAIGFISKATGESVAVG 719

RESULT 6  
 ID 092K07 PRELIMINARY; PRT; 1291 AA.  
 AC 092K07  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE HYPOTHETICAL PROTEIN SMC01708.  
 GN SMC01708.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021.  
 RX MEDLINE=21368234; PubMed=11474104;  
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Bothe G.,  
 Boutry M., Bowser L., Buhmester J., Cadieu E., Capela D., Chain P.,  
 Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,  
 Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,  
 Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
 Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaire V.,  
 Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
 Ramsperger U., Surzycki R., Thebault P., Vandenberg M.,  
 Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,  
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti,"  
 RL Science 293:668-672(2001).  
 DR EMBL; AL591783; CAC41895.1;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 1291 AA; 127509 MW; 1B0F2A8CD1B13C CRC64;

Query Match 32.1%; Score 67; DB 16; Length 1291;  
 Best Local Similarity 44.2%; Pred. No. 20;  
 Matches 19; Conservative 3; Mismatches 13; Indels 8; Gaps 1;  
 OY 4 SEADGKGG-----ANAGDKSIAIGDIAQALGSOSIAIG 38  
 DB 88 NEADPGASDAIAIGTDIAQANDRSIAIGRONAGNEGSIGIG 130  
 ID 09PD63 PRELIMINARY; PRT; 1004 AA.  
 AC 09PD63  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE SURFACE-EXPOSED OUTER MEMBRANE PROTEIN.  
 GN XFI516.  
 OS Xylella fastidiosa.  
 CC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hohnselt J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,  
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 Vaitava H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.T.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 Marques M.V., Marzeca E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Vailida H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 Zago M.A., Zatz M., Zeldis U., Setubal J.C.,  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa,"  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003981; AAF84325.1;  
 KW Complete proteome.  
 SO SEQUENCE 1004 AA; 98283 MW; 0BFF04D0F048B11 CRC64;

Query Match 31.6%; Score 66; DB 16; Length 1004;  
 Best Local Similarity 50.0%; Pred. No. 20;  
 Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
 OY 9 GKGANAGDKSIAIGDIAQALGSOSIAIG 38  
 DB 576 GSGVASAIGDKSTATGSAQAVDSVALG 605  
 ID 09PC04 PRELIMINARY; PRT; 1190 AA.  
 AC 09PC04  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE SURFACE PROTEIN.  
 GN XFI981.  
 OS Xylella fastidiosa.  
 CC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hohnselt J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,  
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 Vaitava H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.T.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE004017; AAF84783.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;

Query Match 30.9%; Score 64.5; DB 16; Length 1190;  
 Best Local Similarity 44.4%; Pred. No. 36;  
 Matches 16; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

OY 4 SEADGKGGAN-ARGDKSIAIGDIAQALGQSIAIG 38  
 DB 191 SKTDCNTSGATVAGLRISIAIGTTARSQSDAISIG 226

RESULT 9

ID 09F2D8 PRELIMINARY; PRT; 1107 AA.  
 AC 09F2D8;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SAPB PROTEIN.  
 GN SAPB.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RA Henderson I.R., Nataro J.P., Cappello R., Stehn C.;  
 RT "Evolutionary origins of the autotransporter proteins."  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ277623; CAC14217.1; -  
 SQ SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;

Query Match 30.6%; Score 64; DB 2; Length 1107;  
 Best Local Similarity 38.5%; Pred. No. 38;  
 Matches 15; Conservative 7; Mismatches 15; Indels 2; Gaps 1;

OY 4 SEADGKGGANARGDKSIAIGDIAQALGQSIAIGDKI 42  
 DB 879 TWTDC--ADANAGADSVAGISGSIAMENSVALTNSV 915

RESULT 10

ID 085267 PRELIMINARY; PRT; 454 AA.  
 AC 085267;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE ADHESIN.  
 GN YADA.  
 OS Yersinia enterocolitica.  
 OC Plasmid pYV, and Plasmid pYve227.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OX NCBI\_TaxID=630;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Iriarte M., Kerbouch C., Lambermont I., Cornelis G.R.;  
 RT "Detailed genetic map of the pYV plasmid of *Yersinia enterocolitica*  
 O:3.";  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Iriarte M., Kerbouch C., Cornelis G.R.;  
 RT "Yada" and ORF291 of *Yersinia enterocolitica* O:3.";

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W22703;  
 RA Iriarte M., Lambermont I., Kerbouch C., Cornelis G.R.;  
 RT "Detailed genetic map of the pYve227 plasmid of *Yersinia*  
 enterocolitica serotype O:3.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF060892; AAC33679.1; -  
 DR EMBL: AF102990; AAD16868.1; -  
 KW Plasmid.  
 SQ SEQUENCE 454 AA; 47270 MW; 2E05D5689EF03014 CRC64;

Query Match 30.4%; Score 63.5; DB 2; Length 454;  
 Best Local Similarity 44.2%; Pred. No. 16;  
 Matches 19; Conservative 5; Mismatches 10; Indels 9; Gaps 2;

OY 9 GKGAN--ARGDKSIAIGDIAQALGQSIAIGDKI 42  
 DB 59 GAGLNCASAKSHISIAIGATTEAKKAAGAVAGSIAIGVNSV 101

RESULT 11

ID 09EQT3 PRELIMINARY; PRT; 261 AA.  
 AC 09EQT3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE GTP-BINDING PROTEIN LIKE 1 (WRCH-1).  
 GN WRCH1 OR MG28K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Delgado Y., Takayama I., Fujino M.A.;  
 RT "Isolation, mapping, and characterization of a novel murine cDNA,  
 RT MG28K homologous to the mammalian GTP-binding protein family.";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21352806; PubMed=11459829;  
 RA Tec W., Pennica D., Xu L., Kaleyta R.F., Levine A.J.;  
 RT "Wrch-1, a novel member of the Rho gene family that is regulated by  
 RT Wnt-1.";  
 RL Genes Dev. 15:1796-1807(2001).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.  
 DR EMBL: AB051827; BAB18639.1; -  
 DR EMBL: AF378088; AAK83541.1; -  
 DR HSP: P21181; IMA4.  
 DR MGD: MG1:1916831; Wrch1.  
 DR InterPro: IPR001806; Ras\_transfmrng.  
 DR InterPro: IPR003578; Rho.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00174; RHO. 1.  
 KW GTP-binding; Lipoprotein; Prenylation.  
 SQ SEQUENCE 261 AA; 28353 MW; BEC4250071E75B14 CRC64;

Query Match 30.1%; Score 63; DB 11; Length 261;  
 Best Local Similarity 38.9%; Pred. No. 9.9;  
 Matches 14; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

OY 2 GISEADGKGGANARGDKSIAIGDIAQALGQSIAIG 37  
 DB 37 GVSAGRGAGAGRGKVCVLD--GAVGKISLIV 70

RESULT 12  
 091A53



ID Q9LA53 PRELIMINARY; PRT; 487 AA.  
 AC Q9LA53;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE IMMUNOGLOBULIN-BINDING PROTEIN E1BE.  
 GN E1BE.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ECOR-9;  
 RX MEDLINE=20187497; PubMed=10722621;  
 RA Sandt C.H., Hill C.W.;  
 RT "Four different genes responsible for nonimmune immunoglobulin-binding  
 activities within a single strain of Escherichia coli."  
 RL Infect. Immun. 68:2205-2214(2000).  
 DR EMBL: AF151676; AAF63045.1;  
 SQ SEQUENCE 487 AA; 51605 MW; D959C67EF7C21CF3 CRC64;  
 Query Match 30.1%; Score 63; DB 2; Length 487;  
 Best Local Similarity 33.3%; Pred. No. 20;  
 Matches 19; Conservative 7; Mismatches 11; Indels 20; Gaps 2;  
 OY 2 GISEADGKG-----GANNRDKSIAIGDIAQ-----ALGSGSIAIG 38  
 DB 138 GPSVEDGKGVSLGSDAKSGEFSVAVNGAKATEKASTVSGWSAALGROSVALG 194  
 RESULT 13  
 Q9PD50 PRELIMINARY; PRT; 2059 AA.  
 AC Q9PD50;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SURFACE PROTEIN.  
 GN XFI329.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 OX NCBI\_Taxid=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 Bueno M.R.P., Camargo A., Camargo L.E.A., Carraro D.M., Carier H.,  
 Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 Gariet M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hohnselt J.D., Junqueira M.U., Kemper E.L., Kitajima J.P.,  
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 Marques M.V., Martins E.A.L., Martins E.M.F., Matukuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 Nhami A.J.R., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Pesquero J.B.,  
 Peloto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 de Rosa V.E.Jr., de Sa R.G., Sancelli R.V., Sawasaki H.E.,  
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,  
 da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 Zago M.A., Zatz M., Zeldanis J., Zetudal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE003982; AAF6438.1;  
 KW Complete proteome.  
 SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5D24E10 CRC64;  
 Query Match 30.1%; Score 63; DB 16; Length 2059;  
 Best Local Similarity 36.7%; Pred. No. 1e+02;  
 Matches 18; Conservative 4; Mismatches 19; Indels 8; Gaps 1;  
 OY 2 GISEADGKG-----GANNRDKSIAIGDIAQALGSGSIAIGDNKI 42  
 DB 1731 GNYDGDATGSKAIAAGVGTQASGEAAVSGGAAGSKSTAIARRAI 1779  
 RESULT 14  
 Q9S2C3 PRELIMINARY; PRT; 436 AA.  
 AC Q9S2C3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PROBABLE SOLUTE-BINDING PROTEIN.  
 GN SCE85.18.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.  
 OX NCBI\_Taxid=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 Kishashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL110470; CAB54175.1;  
 DR InterPro: IPR000567; SBP\_Dbc\_1.  
 DR Pfam: PF01547; SBP\_bacterial\_1;  
 SQ SEQUENCE 436 AA; 46593 MW; 55815BEDDA6C4C04 CRC64;  
 Query Match 29.9%; Score 62.5; DB 2; Length 436;  
 Best Local Similarity 35.8%; Pred. No. 20;  
 Matches 19; Conservative 5; Mismatches 16; Indels 13; Gaps 2;  
 OY 2 GISEADGKG-----KCGA-----NARGDKSIAIGDIAQALGSGSIAIGDNK 41  
 DB 283 GIGADGDGDFVSSGLGVMNPPYDGKGDPSNVAVNPAYLISSTKATDEK 335  
 RESULT 15  
 Q9AJV8 PRELIMINARY; PRT; 550 AA.  
 AC Q9AJV8;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE U... PROBABLE PERIPLASMIC UDP-SUGAR HYDROLASE.

CN [USHA](#).  
 OS *Shigella flexneri*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Shigella*.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M90T;  
 RX MEDLINE=20566792; Pubmed=1111511;  
 RA Buchrieser C., Glaeser P., Rumnick C., Nedjari H., d'Hauteville H.,  
 RA Kunst F., Sansonetti P., Parosot C.;  
 RT "The virulence plasmid pWR100 and the repertoire of proteins secreted  
 by the type III secretion apparatus of *Shigella flexneri*.";  
 RL *Mol. Microbiol.* 38:760-771(2000).  
 DR EMBL; AL391753; CAC05838.1; -.  
 DR HSSP; P07024; 2USH.  
 DR InterPro: IPR001224; 5\_nucleotidase.  
 DR InterPro: IPR000934; Ser\_thr\_phosphase.  
 DR Pfam; PF01009; 5\_nucleotidase; 1.  
 DR PROSITE; PS00785; 5\_NUCLEOTIDASE; 1.  
 DR PROSITE; PS00786; 5\_NUCLEOTIDASE\_2; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 550 AA; 61199 MW; ABA6B2165DD3A83 CRC64;

[illegible]

Search completed: July 30, 2002, 15:37:26  
Job time: 331 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:11 ; Search time 82.85 Seconds  
(without alignments)  
10.725 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41  
Sequence: 1 GTVVLGKK 8

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	8	18	AAW32275
2	41	100.0	8	22	AAE00704
3	36	87.8	7	22	AAE00817
4	36	87.8	1992	17	AAW04505
5	36	87.8	1992	22	AAE69133
6	36	87.8	1992	22	AAE69137
7	36	87.8	2047	22	AAE69134
8	36	87.8	2053	22	AAE69135
9	36	87.8	2123	22	AAE00701
10	36	87.8	2314	22	AAE69136
11	35	85.4	509	18	AAW18786

12	33	80.5	109	22	ABW17879	Human nervous syst
13	33	80.5	184	21	AAE61517	Arabidopsis thalia
14	33	80.5	185	21	AAE61517	Arabidopsis thalia
15	33	80.5	190	21	AAE61516	Arabidopsis thalia
16	33	80.5	191	21	AAE61517	Arabidopsis thalia
17	33	80.5	193	21	AAE61515	Arabidopsis thalia
18	33	80.5	194	21	AAE61516	Arabidopsis thalia
19	33	80.5	537	22	AAE58083	Drosophila melanog
20	33	80.5	545	22	AAE66603	Drosophila melanog
21	32	78.0	174	19	AAE53995	Porcine reproducti
22	32	78.0	372	21	AAE28809	Arabidopsis thalia
23	32	78.0	373	21	AAE28808	Arabidopsis thalia
24	32	78.0	375	21	AAE28807	Arabidopsis thalia
25	32	78.0	470	22	AAE27998	Novel human diagno
26	32	78.0	494	22	AAE41166	Human polypeptide
27	32	78.0	933	20	AAE21621	Ligand binding dom
28	32	78.0	933	21	AAE97297	Human progesterone
29	32	78.0	960	22	AAE58789	Drosophila melanog
30	32	78.0	960	22	AAE66451	Drosophila melanog
31	32	78.0	960	22	AAE66452	Drosophila melanog
32	32	78.0	11096	22	AAE10129	Streptomyces nours
33	31	75.6	68	22	ABE30910	Peptide #3561 enco
34	31	75.6	68	22	ABE36093	Peptide #3599 enco
35	31	75.6	68	22	ABE21484	Protein #3483 enco
36	31	75.6	68	22	AAE56875	Human brain expres
37	31	75.6	68	22	AAE69262	Human bone marrow
38	31	75.6	68	22	AAE17094	Peptide #3528 enco
39	31	75.6	68	22	AAE29584	Peptide #3474 enco
40	31	75.6	68	22	AAE04792	Human polyprotein
41	31	75.6	75	22	AAE13037	Propionibacterium
42	31	75.6	86	22	AAE67949	Human prostate can
43	31	75.6	167	22	AAE63802	H. pylori ORF 029e
44	31	75.6	220	19	AAE11027	H. pylori ORF 635
45	31	75.6	220	19	AAE98208	H. pylori ORF 635

## ALIGNMENTS

RESULT 1  
AAW32275 standard; peptide; 8 AA.

AC AAW32275;

DT 08-MAY-1998 (first entry)

DE M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 2.

DE Outer membrane protein-106; OMP106; vaccine; immune response;

KW cytotoxic antibody; Moraxella catarrhalis.

OS Moraxella catarrhalis.

PN W09741731-AL.

PD 13-NOV-1997.

PE 28-APR-1997; 97WO-US07679.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Plosila L, Tucker K;

DR WPI; 1997-558601/51.

PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in

PS vaccines for producing immune responses against M. catarrhalis

PS Claim 10; Page 23; 78pp; English.

CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)  
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella  
 CC catarhalis, an haemagglutinating cultivar. The peptide fragment can  
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The  
 CC antibody is a cytotoxic antibody which mediates complement killing of  
 CC M. catarhalis. The OMP106 polypeptide, and its peptide fragments can be  
 CC used in vaccines and antigenic compositions. They can also be used for  
 CC producing an immune response in an animal against M. catarhalis.  
 XX  
 SQ Sequence 8 AA;

Query Match 100.0%; Score 41; DB 18; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYVLGKK 8  
 DB 1 gvlvlgkk 8

## RESULT 2

ID AAE00704 standard; peptide; 8 AA.

AC AAE00704;

DT 02-JUL-2001 (first entry)

DE Internal peptide of M. catarhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antidiabetic;

OS Moraxella catarhalis.

PN US6214981-B1.

PD 10-APR-2001.

PE 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarhalis outer  
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial  
 PT infections and as vaccine against Moraxella catarhalis infection of  
 PT mammals -

PS Example: Column 31; 49pp; English.

CC The present sequence is internal peptide of haemagglutinating  
 CC Moraxella catarhalis outer membrane protein-106 (OMP106). The OMP106 is  
 CC used as a therapeutic and prophylactic vaccine against M. catarhalis  
 CC infections of mammals. It is used for diagnosis of bacterial infections  
 CC and as reagents for clinical or medical diagnosis of M. catarhalis  
 CC infections and for scientific research on the properties of  
 CC pathogenicity, virulence and infectivity of M. catarhalis. It is also  
 CC used as a probe to identify the presence of M. catarhalis in biological  
 CC specimens and to identify other bacteria that encode a polypeptide  
 CC related to M. catarhalis OMP106. OMP106-derived polypeptides are used  
 CC as ligands to detect antibodies elicited in response to M. catarhalis  
 CC infections and also as immunogens for inducing M. catarhalis-specific  
 CC antibodies which are useful in immunoassays to detect M. catarhalis in  
 CC biological specimens. Cytotoxic antibodies are useful in passive  
 CC immunisations against M. catarhalis.

XX Sequence 8 AA;

SQ Query Match 100.0%; Score 41; DB 22; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYVLGKK 8  
 DB 1 gvlvlgkk 8

## RESULT 3

ID AAE00817 standard; peptide; 7 AA.

AC AAE00817;

DT 02-JUL-2001 (first entry)

DE Peptide related to the invention.

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antidiabetic;

OS Unidentified.

PN US6214981-B1.

PD 10-APR-2001.

PE 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarhalis outer  
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial  
 PT infections and as vaccine against Moraxella catarhalis infection of  
 PT mammals -

PS Disclosure: Column 63; 49pp; English.

CC The present sequence is a peptide which is related to the invention.  
 CC The present invention relates to haemagglutinating  
 CC Moraxella catarhalis outer membrane protein-106 (OMP106). The OMP106 is  
 CC used as a therapeutic and prophylactic vaccine against M. catarhalis  
 CC infections of mammals. It is used for diagnosis of bacterial infections  
 CC and as reagents for clinical or medical diagnosis of M. catarhalis  
 CC infections and for scientific research on the properties of  
 CC pathogenicity, virulence and infectivity of M. catarhalis. It is also  
 CC used as a probe to identify the presence of M. catarhalis in biological  
 CC specimens and to identify other bacteria that encode a polypeptide  
 CC related to M. catarhalis OMP106. OMP106-derived polypeptides are used  
 CC as ligands to detect antibodies elicited in response to M. catarhalis  
 CC infections and also as immunogens for inducing M. catarhalis-specific  
 CC antibodies which are useful in immunoassays to detect M. catarhalis in  
 CC biological specimens. Cytotoxic antibodies are useful in passive  
 CC immunisations against M. catarhalis.  
 CC Note: No other information is provided for this sequence in the  
 CC specification.

SQ Sequence 7 AA;

Query Match 87.8%; Score 36; DB 22; Length 7;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGK 7  
111111  
Db 1 gtvlgk 7

## RESULT 4

AAW04505 standard; Protein; 1992 AA.

AAW04505;

25-JAN-1997 (first entry)

Moraxella 200 kDa outer membrane protein.

Outer membrane protein; OMP; immunogen; vaccine; otitis media;

diagnosis.

Mycobacterium catarrhalis strain 4223.

MO9634960-A1

07-NOV-1996.

29-APR-1996; 96WO-CA00264.

26-MAR-1996; 96US-0621944.

01-MAY-1995; 95US-0431718.

07-JUN-1995; 95US-0478370.

(CONN-) CONNAUGHT LAB LTD.

Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;

WPI: 1996-506162/50.

N-PSDB; AAT38740.

Moraxella outer membrane protein - useful as immunogen in protective

vaccine and for diagnosis

Claim 14; Fig 6; 109pp; English.

An approx. 200 kDa outer membrane protein (AAW04505) can be

isolated from Moraxella catarrhalis otitis media strain 4223

by electroporation, or expressed from a gene (see also AAT38740)

obtd. from a strain 4223 genomic library. Natural or recombinant

outer membrane protein is useful as an immunogen to protect

against infection by Moraxella, esp. M. catarrhalis. It can

also be used to detect antibodies, esp. for differential diagnosis

between bacteria that cause similar symptoms; and also useful as

a carrier for other antigens and used to raise antitumour

antibodies for conjugation to therapeutic agents.

Sequence 1992 AA;

Query Match 87.8%; Score 36; DB 17; Length 1992;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 24-APR-2001 (first entry)

M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

Moraxella catarrhalis strain 4223; major outer membrane protein;

200kDa outer membrane protein; antibacterial; immunogenic; infection;

otitis media; detection.

Moraxella catarrhalis.

MO200107619-A1.

01-FEB-2001.

26-JUL-2000; 2000WO-CA00870.

27-JUL-1999; 99US-0361619.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Sasaki K, Yang Y, Klein MH;

WPI: 2001-159722/16.

N-PSDB; AAF59100, AAF59101.

New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

useful in protective vaccines and for diagnosis

Example 3; Fig 2A-W; 247pp; English.

The present invention describes an isolated and purified nucleic acid (I)

that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

The 200 kDa outer membrane protein (II) has antibacterial activity and

can be used in vaccines. (II), and its truncated versions, are used as

immunogenic compositions and vaccines to protect against M. catarrhalis

infections, particularly otitis media in humans. (II) is also used as

antigen in immunoassays for detecting specific antibodies (Ab), and to

generate Ab. (I) are used for recombinant production of (II) and its

fragments are used as probes for identifying/cloning 200 kDa protein

genes from other strains, and for diagnostic detection of M. catarrhalis.

(I) makes possible production of large amount of recombinant immunogens.

Expression of truncated versions of (II) reduces toxicity of the protein

towards the Escherichia coli host. The present sequence represents the

M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is

used in the exemplification of the present invention.

Sequence 1992 AA;

Query Match 87.8%; Score 36; DB 22; Length 1992;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGK 7  
111111  
Db 1604 gtvlgk 1610

RESULT 6

AAW04505 standard; Protein; 1992 AA.

AAW04505;

24-APR-2001 (first entry)

M. catarrhalis M56 200kDa protein in PK348 SEQ ID NO:13.

Moraxella catarrhalis strain 08; major outer membrane protein;

200kDa outer membrane protein; antibacterial; immunogenic; infection;

otitis media; detection.

XX PA (CONN-) CONNAUGHT LAB LTD.  
XX  
XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
XX DR WPI; 2001-159722/16.  
XX DR N-PSTDB; AAF59102, AAF59103.  
  
New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
useful in protective vaccines and for diagnosis

Claim 1; Fig 3A-W; 247pp; English.

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (III), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (IV) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (V) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (VI) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 genomic 200kda protein, which is given in the exemplification of the present invention.

SQ Sequence      2047 AA;

Query Match                  87.8%; Score 36; DB 22; Length 2047;  
Best Local Similarity    100.0%; Pred. No. 4,1e+02;  
Matches         7; Conservative        0; Mismatches        0; Indels        0; Gaps        0

OY        1 GTVLGK 7  
         :|||  
Db       1659 gtlvlgk 1665

	RESULT	8
AAB69135		
ID	AAB69135 standard; protein; 2053 AA.	
AC	AAB69135;	
XX		
DE	24-APR-2001 (first entry)	
DT		
XX	M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.	
XX		
KW	Moraxella catarrhalis strain Q8; major outer membrane protein;	
KW	200kDa outer membrane protein; antibacterial; immunogenic; infection;	
KW	otitis media; detection.	
XX		
OS	Moraxella catarrhalis.	
XX		
WO	200107619-A1.	
PN		
PD	01-FEB-2001.	
XX		
PF	26-JUL-2000; 2000WO-CAN00870.	
XX		
PR	27-JUL-1999; 99US-0361619.	
PA	(CONN-) CONNADUGHT LAB LTD.	
XX		
XX	Locsmore SM, Sasaki K, Yang Y, Klein MH;	
PI		
DR	WPI: 2001-159722/16.	
DR	N-FSDS; AAF59104.	
XX		
XT	New nucleic acid encoding Moraxella catarrhalis outer membrane protein	

PT	useful in protective vaccines and for diagnosis.
XX	
PS	Claim 1; Fig 4A-V; 247pp; English.
CC	The present invention describes an isolated and purified nucleic acid (I)
CC	that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC	The 200 kDa outer membrane protein (II) has antibacterial activity and
CC	can be used in vaccines. (II), and its truncated versions, are used as
CC	immunogenic compositions and vaccines to protect against M. catarrhalis
CC	infections, particularly otitis media in humans. (II) is also used as
CC	antigen in immunoassays for detecting specific antibodies (Ab), and to
CC	generate Ab. (I) are used for recombinant production of (II) and its
CC	fragments are used as probes for identifying/cloning 200 kDa protein
CC	genes from other strains, and for diagnostic detection of M. catarrhalis.
CC	(I) makes possible production of large amount of recombinant immunogens.
CC	Expression of truncated versions of (II) reduces toxicity of the protein
CC	towards the Escherichia coli host. The present sequence represents the
CC	M. catarrhalis strain 08 200kDa protein, which is given in the
CC	exemplification of the present invention.
XX	
SQ	Sequence 2053 AA;
XX	
Query Match	87.8%; Score 36; DB 22; Length 2053;
Best Local Similarity	100.0%; Pred. No. 4,1e+02;
Matches 7; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 GTVLGSK 7
Db	1665 gtvlgsk 1671
RESULT 9	
AAE00701	
ID AAE00701 standard; Protein; 2123 AA.	
XX	
AC AAE00701;	
XX	
DT 02-JUL-2001 (first entry).	
XX	
DE Moraxella catarrhalis outer membrane protein-106 (OMP106).	
XX	
KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;	
KM bacterial infection; Immunogen; cytotoxic; antibiotic;	
KW passive immunisation.	
OS Moraxella catarrhalis.	
XX	
PN US6214981-B1.	
XX	
PD 10-APR-2001.	
XX	
PF 12-NOV-1997; 97US-0968685.	
XX	
PR 03-MAY-1996; 96US-0642712.	
XX	
PA (ANTE-) ANTEX BIOLOGICS INC.	
XX	
PI Tucker K, Plosila L, Tillman UF;	
XX	
DR WPI; 2001-281002/29.	
XX	
DR N-PSTDB; AAD04029.	
XX	
PT Novel nucleotide sequences encoding Moraxella catarrhalis outer	
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial	
PT infections and as vaccine against Moraxella catarrhalis infection of	
PT mammals -	
XX	
PS Claim 7; Column 53-64; 49pp; English.	
XX	
CC The present sequence is haemagglutinating Moraxella catarrhalis outer	
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic	
and prophylactic vaccine against M. catarrhalis infections of mammals.	

CC	It is used for diagnosis of bacterial infections and as reagents for
CC	clinical or medical diagnosis of M. catarrhalis infections and for
CC	scientific research on the properties of pathogenicity, virulence and
CC	infectivity of M. catarrhalis. It is also used as a probe to identify
CC	the presence of M. catarrhalis in biological specimens and to identify
CC	other bacteria that encode a polypeptide related to M. catarrhalis
CC	OMP106. OMP106-derived polypeptides are used as ligands to detect
CC	antibodies elicited in response to M. catarrhalis infections and also
CC	as immunogens for inducing M. catarrhalis-specific antibodies which are
CC	useful in immunoassays to detect M. catarrhalis in biological specimens.
CC	Cytotoxic antibodies are useful in passive immunisations against
CC	M. catarrhalis.
XX	
SO	Sequence 2123 AA:
Query Match	87.8%; Score 36; DB 22; Length 2123;
Best Local Similarity	100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Oy	1 GTVLGK 7 
Db	1735 gtvlgk 1741
RESULT 10	
AAB69136	
ID	AAB69136 standard; Protein; 2314 AA.
XX	
AC	AAB69136;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	M. catarrhalis les1 200kDa protein SEQ ID NO:11.
XX	
KM	Moraxella catarrhalis strain 08; major outer membrane protein;
KM	200kDa outer membrane protein; antibacterial; immunogenic; infection;
KM	otitis media; detection.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO200107619-A1.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-COA0870.
XX	
PR	27-JUL-1999; 99US-0361619.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX	
DR	WPI: 2001-159722/16.
XX	
PT	N-TSDB; AAF59105.
XX	
PS	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX	useful in protective vaccines and for diagnosis
Claim 1; Fig 5A-Y; 247pp; English.	
XX	The present invention describes an isolated and purified nucleic acid (I)
CC	that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC	The 200 kDa outer membrane protein (II) has antibacterial activity and
CC	can be used in vaccines. (II), and its truncated versions, are used as
CC	immunogenic compositions and vaccines to protect against M. catarrhalis
CC	infections, particularly otitis media in humans. (II) is also used as
CC	antigen in immunoassays for detecting specific antibodies (Ab), and to
CC	generate Ab. (I) are used for recombinant production of (II) and its
CC	fragments are used as probes for identifying/cloning 200 kDa protein
CC	genes from other strains, and for diagnostic detection of M. catarrhalis.
CC	(I) makes possible production of large amount of recombinant immunogens.
CC	Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the Escherichia coli host. The present sequence represents the  
CC M. catarrhalis leel 200kDa protein, which is given in the exemplification  
CC of the present invention.

XX Sequence 2314 AA;

Query Match 87.8%; Score 36; DB 22; Length 2314;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGSK 7  
1111111  
Db 1926 gtlvlgsk 1932

RESULT 11

AAW18786  
ID AAW18786 standard; Protein; 509 AA.

XX AAW18786;

XX 10-DEC-1997 (first entry)

DE Neisseria adhesion protein (from OrfB).

KW Neisseria gonorrhoea; adhesion; Lipoprotein; OrfA; OrfI; OrfB.

OS Neisseria gonorrhoeae.

XX DE19534579-A1.

XX 20-MAR-1997.

XX 18-SEP-1995; 95DE-1034579.

XX 18-SEP-1995; 95DE-1034579.

PA (PLAC) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.

PI Fischer E, Maier J, Meyer TF, Rudel T, Scheuerpflug I;

DR WPL; 1997-180942/17.

PT Nucleic acids encoding Neisseria adhesion proteins - for therapeutic  
PS and diagnostic use

XX Claim 40; Page 11-13; 20pp; German.

CC OrfA and OrfB in complexes with the protein PilC are capable of  
CC adhering to human cells.

CC Products obtained from the DNA are useful in medicaments,  
CC diagnostic compns. and vaccines, esp. for treatment of

CC Neisseria gonorrhoea and N. meningitidis infections.

XX Sequence 509 AA;

Query Match 85.4%; Score 35; DB 18; Length 509;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8  
1111111  
Db 326 gtlvlgsk 333

RESULT 12

ABBI7879  
ID ABBI7879 standard; Protein; 109 AA.

XX AC ABBI7879;

XX

DT 23-JAN-2002 (first entry)

XX Human nervous system related polypeptide SEQ ID NO 6536.

DE Human; nototropic; neuroprotective; cytostatic; dermatological; virucide;

KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;

KW antiparkinsonian; antiscickling; antianemic; antithrillitic; cancer;

KW antipneumatic; hepatotropic; neuroprotective; antiinflammatory;

KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;

KW antipruritic; cardiac; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 16-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226688.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 03-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0230437.

XX 06-SEP-2000; 2000US-0230438.

XX 06-SEP-2000; 2000US-0231242.

XX 08-SEP-2000; 2000US-0231243.

XX 08-SEP-2000; 2000US-0231244.

XX 08-SEP-2000; 2000US-0231413.

XX 08-SEP-2000; 2000US-0231414.

XX 08-SEP-2000; 2000US-0232080.

XX 08-SEP-2000; 2000US-0232081.

XX 12-SEP-2000; 2000US-0231968.

XX 14-SEP-2000; 2000US-0232397.



PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232403.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234224.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 20-OCT-2000; 2000US-0242221.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 01-DEC-2000; 2000US-0251160.  
 PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2000US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI, 2001-541565/60.  
 DR N-PSDB; ABA14205.  
 XX  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 PT cancers and metastases  
 PT  
 XX  
 PS Claim 11: SEQ ID NO 6536; 1701bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcl.sequences.  
 XX  
 SO Sequence 109 AA;  
 OY 1 GTVVGKK 8  
 Db 5 gsvlggqk 12  
 RESULT 13  
 AAG61517  
 ID AAG61517 standard; Protein: 184 AA.  
 XX  
 AC AAG61517;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 79798.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 OS  
 XX EPI033405-A2.  
 FN  
 XX

Query Match 80.5%; Score 33; DB 22; Length 109;  
 Best Local Similarity 75.0%; Pred. No. 68;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129848.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130444.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139465.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 23-OCT-1999; 9905-0161404.  
PR 23-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161359.  
PR 25-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 26-OCT-1999; 9905-0161362.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

Query Match 80.58; Score 33; DB 21; Length 184;  
Best Local Similarity 87.58; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
Db 25 gtlvlgkk 32

## RESULT 14

AAG57078  
ID AAG57078 standard; Protein: 185 AA.

XX AAG57078;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 73506.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.  
PR 09-MAR-1999; 9905-0123548.

PR 23-MAR-1999; 9905-0125788.  
PR 25-MAR-1999; 9905-0126264.  
PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.  
PR 08-APR-1999; 9905-0128714.  
PR 16-APR-1999; 9905-0129845.  
PR 19-APR-1999; 9905-0130077.  
PR 21-APR-1999; 9905-0130449.  
PR 23-APR-1999; 9905-0130510.  
PR 23-APR-1999; 9905-0130891.  
PR 28-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132048.  
PR 04-MAY-1999; 9905-0132407.  
PR 05-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 06-MAY-1999; 9905-0132487.  
PR 07-MAY-1999; 9905-0132863.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135553.  
PR 24-MAY-1999; 9905-0136229.  
PR 25-MAY-1999; 9905-0136621.  
PR 27-MAY-1999; 9905-0136782.  
PR 28-MAY-1999; 9905-0137228.  
PR 01-JUN-1999; 9905-0137228.  
PR 03-JUN-1999; 9905-0137502.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142927.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.

PR 16-JUL-1999; 99US-0144086.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0147935.  
 PR 11-AUG-1999; 99US-0148171.  
 PR 12-AUG-1999; 99US-0148319.  
 PR 13-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159283.  
 PR 13-OCT-1999; 99US-0159284.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 14-OCT-1999; 99US-0159684.  
 PR 18-OCT-1999; 99US-0159884.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160771.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 26-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 185;  
 Best local Similarity 87.5%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STVLGKK 8  
 Db 25 stvlgk 32

RESULT 15  
 ID AAG61516 standard; Protein; 190 AA.  
 XX AAG61516;  
 AC AAG61516;  
 XX 18-OCT-2000 (first entry)  
 DT 18-OCT-2000 (first entry)  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 79797.  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 79797.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX EPI033405-A2.  
 PN EPI033405-A2.  
 XX 06-SEP-2000.  
 PD 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123348.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.

PR	1-APR-1999,	99US-0130077,
PR	21-APR-1999,	99US-0130449,
PR	23-APR-1999,	99US-0130510,
PR	28-APR-1999,	99US-0130861,
PR	28-APR-1999,	99US-0131449,
PR	30-APR-1999,	99US-0132048,
PR	30-APR-1999,	99US-0132407,
PR	04-MAY-1999,	99US-0132484,
PR	05-MAY-1999,	99US-0132485,
PR	06-MAY-1999,	99US-0132486,
PR	07-MAY-1999,	99US-0132487,
PR	11-MAY-1999,	99US-0132863,
PR	11-MAY-1999,	99US-0134256,
PR	14-MAY-1999,	99US-0134218,
PR	14-MAY-1999,	99US-0134219,
PR	14-MAY-1999,	99US-0134321,
PR	14-MAY-1999,	99US-0134370,
PR	14-MAY-1999,	99US-0134768,
PR	19-MAY-1999,	99US-0134941,
PR	20-MAY-1999,	99US-0135124,
PR	21-MAY-1999,	99US-0135353,
PR	24-MAY-1999,	99US-0135629,
PR	25-MAY-1999,	99US-0136021,
PR	27-MAY-1999,	99US-0136382,
PR	28-MAY-1999,	99US-0136782,
PR	01-JUN-1999,	99US-0137222,
PR	03-JUN-1999,	99US-0137528,
PR	04-JUN-1999,	99US-0137502,
PR	07-JUN-1999,	99US-0137724,
PR	08-JUN-1999,	99US-0138094,
PR	10-JUN-1999,	99US-0138540,
PR	10-JUN-1999,	99US-0138847,
PR	14-JUN-1999,	99US-0139119,
PR	16-JUN-1999,	99US-0139452,
PR	16-JUN-1999,	99US-0139453,
PR	17-JUN-1999,	99US-0139452,
PR	18-JUN-1999,	99US-0139454,
PR	18-JUN-1999,	99US-0139455,
PR	18-JUN-1999,	99US-0139456,
PR	18-JUN-1999,	99US-0139457,
PR	18-JUN-1999,	99US-0139458,
PR	18-JUN-1999,	99US-0139459,
PR	18-JUN-1999,	99US-0139460,
PR	18-JUN-1999,	99US-0139461,
PR	18-JUN-1999,	99US-0139462,
PR	18-JUN-1999,	99US-0139463,
PR	18-JUN-1999,	99US-0139750,
PR	18-JUN-1999,	99US-0139763,
PR	21-JUN-1999,	99US-0139817,
PR	22-JUN-1999,	99US-0139899,
PR	23-JUN-1999,	99US-0140035,
PR	23-JUN-1999,	99US-0140034,
PR	24-JUN-1999,	99US-0140699,
PR	24-JUN-1999,	99US-0140823,
PR	29-JUN-1999,	99US-0140991,
PR	30-JUN-1999,	99US-0141287,
PR	01-JUL-1999,	99US-0141842,
PR	01-JUL-1999,	99US-0142154,
PR	02-JUL-1999,	99US-0142055,
PR	06-JUL-1999,	99US-0142390,
PR	08-JUL-1999,	99US-0142803,
PR	09-JUL-1999,	99US-0142920,
PR	12-JUL-1999,	99US-0142977,
PR	13-JUL-1999,	99US-0143542,
PR	14-JUL-1999,	99US-0143624,
PR	15-JUL-1999,	99US-0144005,
PR	16-JUL-1999,	99US-0144085,
PR	16-JUL-1999,	99US-0144086,
PR	19-JUL-1999,	99US-0144325,
PR	19-JUL-1999,	99US-0144331,
PR	19-JUL-1999,	99US-0144332,
PR	19-JUL-1999,	99US-0144333,
PR	19-JUL-1999,	99US-0144334,

PR	9-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144332;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145088;
PR	21-JUL-1999;	9905-0145085;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145192;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145224;
PR	23-JUL-1999;	9905-0145276;
PR	26-JUL-1999;	9905-0145913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147036;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147462;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147935;
PR	10-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148319;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149366;
PR	17-AUG-1999;	9905-0149175;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149723;
PR	20-AUG-1999;	9905-0149723;
PR	20-AUG-1999;	9905-0149929;
PR	20-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;
PR	25-AUG-1999;	9905-0150586;
PR	26-AUG-1999;	9905-0150584;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151303;
PR	31-AUG-1999;	9905-0151438;
PR	01-SEP-1999;	9905-0151930;
PR	07-SEP-1999;	9905-0155363;
PR	10-SEP-1999;	9905-0155070;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154039;
PR	20-SEP-1999;	9905-0154779;
PR	22-SEP-1999;	9905-0155133;
PR	23-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156566;
PR	04-OCT-1999;	9905-0157753;
PR	05-OCT-1999;	9905-0157713;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158232;
PR	12-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0159293;
PR	13-OCT-1999;	9905-0159294;
PR	14-OCT-1999;	9905-0159329;

PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159338.  
PR 18-OCT-1999; 99US-0159384.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 190;  
Best Local Similarity 87.5%; Pred. NO. 1.2e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGCK 8  
DB 31 gtlgqkk 38

Search completed: July 30, 2002, 15:34:12  
Job time: 407 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:51 ; Search time 33.83 seconds  
(without alignments)  
5.776 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGKK 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters:

231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	8	4 US-08-968-685A-2	Sequence 2, Appl
2	36	87.8	7	4 US-08-968-685A-11	Sequence 11, Appl
3	36	87.8	2048	4 US-09-268-347-48	Sequence 48, Appl
4	36	87.8	2123	4 US-08-968-685A-10	Sequence 10, Appl
5	36	87.8	2314	4 US-08-268-347-49	Sequence 49, Appl
6	32	78.0	174	4 US-08-686-968C-10	Sequence 10, Appl
7	32	78.0	363	6 5223606-6	Patent No. 5223606
8	32	78.0	933	4 US-08-764-870-14	Sequence 14, Appl
9	32	78.0	933	4 US-08-980-115-14	Sequence 14, Appl
10	31	75.6	134	4 US-09-367-953B-120	Sequence 120, Appl
11	31	75.6	1144	3 US-08-726-214-6	Sequence 6, Appl
12	30	73.2	20	2 US-08-909-433A-11	Sequence 11, Appl
13	30	73.2	174	5 PCT-US95-10904-51	Sequence 51, Appl
14	30	73.2	282	4 US-09-120-365-79	Sequence 79, Appl
15	30	73.2	282	4 US-09-515-039-79	Sequence 79, Appl
16	30	73.2	339	1 US-08-208-007A-11	Sequence 11, Appl
17	30	73.2	339	1 US-08-330-121B-7	Sequence 7, Appl
18	30	73.2	339	4 US-08-860-255A-7	Sequence 7, Appl
19	30	73.2	339	5 PCT-US95-13820-7	Sequence 7, Appl
20	30	73.2	360	1 US-08-361-708-4	Sequence 4, Appl
21	30	73.2	360	1 US-08-536-277-4	Sequence 4, Appl
22	30	73.2	361	1 US-08-361-708-3	Sequence 3, Appl
23	30	73.2	361	1 US-08-314-309A-25	Sequence 25, Appl
24	30	73.2	361	1 US-08-536-277-3	Sequence 3, Appl
25	30	73.2	822	2 US-08-474-067-9	Sequence 9, Appl
26	30	73.2	822	2 US-08-474-068A-9	Sequence 9, Appl
27	30	73.2	822	2 US-08-472-481-8	Sequence 8, Appl

28	30	73.2	1313	2 US-08-244-537-2	Sequence 2, Appl
29	30	73.2	2325	3 US-08-417-089-6	Sequence 6, Appl
30	30	73.2	2325	4 US-08-695-651-6	Sequence 6, Appl
31	30	73.2	2325	4 US-08-930-285-6	Sequence 6, Appl
32	30	73.2	2325	4 US-08-695-421-6	Sequence 6, Appl
33	29	70.7	19	4 US-09-025-596-8	Sequence 8, Appl
34	29	70.7	72	1 US-08-253-155A-48	Sequence 48, Appl
35	29	70.7	108	4 US-09-240-274-51	Sequence 51, Appl
36	29	70.7	108	4 US-09-240-274-52	Sequence 52, Appl
37	29	70.7	108	4 US-09-240-274-53	Sequence 53, Appl
38	29	70.7	108	4 US-09-240-274-166	Sequence 166, Appl
39	29	70.7	1276	1 US-08-222-616-24	Sequence 24, Appl
40	29	70.7	1276	5 PCT-US95-04228-24	Sequence 24, Appl
41	29	70.7	1580	2 US-08-804-227C-11	Sequence 11, Appl
42	29	70.7	1580	2 US-08-804-198-5	Sequence 5, Appl
43	29	70.7	1891	2 US-08-804-227C-12	Sequence 12, Appl
44	29	70.7	1891	2 US-08-804-198-6	Sequence 6, Appl
45	28	68.3	163	1 US-08-044-618-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-968-685A-2  
Sequence 2, Application US/08968685A  
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH

APPLICANT: PLOSILA, LAURA

TITLE OF INVENTION: MORAXELLA CAVARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A

FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66144 PENNIE

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

FRAGMENT TYPE: Internal

US-08-968-685A-2

Query Match 100.0%; Score 41; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches: 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTVLGKK 8

DB 1 GTVLGSK 8

RESULT 2

US-08-968-685A-11  
Sequence 11, Application US/08968685A  
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSTIA, LAURA

TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

FRAGMENT TYPE: Internal

US-08-968-685A-11

Query Match

Best Local Similarity 87.8%; Score 36; DB 4; Length 7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7

DB 1 GTVLGSK 7

RESULT 3

US-09-268-347-48

Sequence 48, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.  
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860

CURRENT APPLICATION NUMBER: US/09/268,347

NUMBER OF SEQ ID NOS: 54

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 48

LENGTH: 2048

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-268-347-48

Query Match

Best Local Similarity 87.8%; Score 36; DB 4; Length 2048;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7

DB 1660 GTVLGSK 1666

RESULT 4

US-08-968-685A-10

Sequence 10, Application US/08968685A

Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSTIA, LAURA

TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2123 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-968-685A-10

Query Match

Best Local Similarity 87.8%; Score 36; DB 4; Length 2123;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7

DB 1735 GTVLGSK 1741

RESULT 5

US-09-268-347-49

Sequence 49, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.  
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS



FILE REFERENCE: 1038-860  
CURRENT APPLICATION NUMBER: US/09/268,347  
CURRENT FILING DATE: 1999-03-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 49  
LENGTH: 2314  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-268-347-49

Query Match 87.8%; Score 36; DB 4; Length 2314;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7  
|||||  
DB 1926 GTVLGK 1932

RESULT 6  
US-08-686-968C-10  
Sequence 10, Application US/08686968C  
Patent No. 6221361  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
APPLICANT: Junker, David E.  
TITLE OF INVENTION: Recombinant Swinepox Virus  
FILE REFERENCE: 39119-H/JML  
CURRENT APPLICATION NUMBER: US/08/686,968C  
CURRENT FILING DATE: 1996-07-25  
NUMBER OF SEQ ID NOS: 231  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 174  
TYPE: PRT  
ORGANISM: Porcine reproductive and respiratory syndrome virus  
US-08-686-968C-10

Query Match 78.0%; Score 32; DB 4; Length 174;  
Best Local Similarity 75.0%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
|||||  
DB 153 GTVLGK 160

RESULT 7  
5223606-6  
Patent No. 5223606  
APPLICANT: BLAUDIN DE THE, HUGHES; MARCHIO, AGNES; TIOLLAIS,  
PIERRE-DEJAN, ANNE  
TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED  
PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA  
NUMBER OF SEQUENCES: 11  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/134,130  
FILING DATE: 17-DEC-1987  
PRIOR APPLICATION DATA:  
SEQ ID NO: 6;  
LENGTH: 363  
5223606-6

Query Match 78.0%; Score 32; DB 6; Length 363;  
Best Local Similarity 75.0%; Pred. No. 89;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
|||||

DB 65 GMLVGRK 72

RESULT 8  
US-08-764-870-14  
Sequence 14, Application US/08764870  
Patent No. 6236946  
GENERAL INFORMATION:  
APPLICANT: Scanlan, Thomas S  
APPLICANT: Baxter, John D  
APPLICANT: Fletterick, Robert J  
APPLICANT: Wagner, Richard L  
APPLICANT: Kushner, Peter J  
APPLICANT: Apriletti, James W  
APPLICANT: West, Brian  
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Cooley Godward  
STREET: Five Palo Alto Square, 3000 El Camino Real  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,870  
FILING DATE: 13-DEC-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,540  
FILING DATE: 13-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,543  
FILING DATE: 13-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,606  
FILING DATE: 14-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Jackie N  
REGISTRATION NUMBER: 35,966  
REFERENCE/DOCKET NUMBER: UCAL-246/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)843-5000  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 933 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-764-870-14

Query Match 78.0%; Score 32; DB 4; Length 933;  
Best Local Similarity 75.0%; Pred. No. 2,4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
|||||  
DB 631 GMLVGRK 638

RESULT 9  
US-08-980-115-14  
Sequence 14, Application US/08980115  
Patent No. 626622  
GENERAL INFORMATION:

APPLICANT: Scanlan, Thomas S.  
APPLICANT: Baxter, John D.  
APPLICANT: Fletchick, Robert J.  
APPLICANT: Wagner, Richard L.  
APPLICANT: Kushner, Peter J.  
APPLICANT: Apiletti, James W.  
APPLICANT: West, Brian L.  
APPLICANT: Shian, Andrew K.  
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS  
FILE REFERENCE: UCAL-246/0205  
CURRENT APPLICATION NUMBER: US/08/980,115  
CURRENT FILING DATE: 1997-11-26  
EARLIER APPLICATION NUMBER: 08/764,870  
EARLIER FILING DATE: 1996-12-13  
EARLIER APPLICATION NUMBER: 60/008,606  
EARLIER FILING DATE: 1995-12-14  
EARLIER APPLICATION NUMBER: 60/008,543  
EARLIER FILING DATE: 1995-12-13  
EARLIER APPLICATION NUMBER: 60/008,540  
EARLIER FILING DATE: 1995-12-13  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 933  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (659)..(918)  
OTHER INFORMATION: minimal ligand binding domain  
US-08-980-115-14

Query Match 78.0%; Score 32; DB 4; Length 933;  
Best Local Similarity 75.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8  
Db 631 GTVLGSK 638

RESULT 10  
US-09-367-953B-120  
Sequence 120, Application US/09367953B  
Patent No. 6287572  
GENERAL INFORMATION:  
APPLICANT: Kingsman, Alan J  
APPLICANT: Kingsman, Susan M  
TITLE OF INVENTION: Anti-HIV peptides and proteins  
FILE REFERENCE: 550-154  
CURRENT APPLICATION NUMBER: US/09/367,953B  
CURRENT FILING DATE: 1999-08-24  
PRIOR APPLICATION NUMBER: PCT/GB98/00563  
PRIOR FILING DATE: 1998-02-23  
PRIOR APPLICATION NUMBER: GB 9703802.0  
PRIOR FILING DATE: 1997-02-24  
NUMBER OF SEQ ID NOS: 131  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 120  
LENGTH: 134  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus type 2  
US-09-367-953B-120

Query Match 75.6%; Score 31; DB 4; Length 134;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8  
Db 5 GTVLGSK 12

RESULT 11  
US-08-726-214-6  
Sequence 6, Application US/08726214  
Patent No. 6107076  
GENERAL INFORMATION:  
APPLICANT: Tang, Wei-Jen  
APPLICANT: Gilman, Alfred G.  
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE  
TITLE OF INVENTION: AND USES THEREFOR  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,214  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,498  
FILING DATE: 04-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:450  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1144 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-726-214-6

Query Match 75.6%; Score 31; DB 3; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLG 6  
Db 418 GTVLG 423

RESULT 12  
US-08-909-433A-1  
Sequence 1, Application US/08909433A  
Patent No. 5981698  
GENERAL INFORMATION:  
APPLICANT: Andrew Brittain  
TITLE OF INVENTION: ACTIVE FRAGMENT OF  
TITLE OF INVENTION: MODIFIED INSECTICIDAL CRYSTAL PROTEINS OF  
TITLE OF INVENTION: BACILLUS THURINGIENSIS-RELATED POLY-  
TITLE OF INVENTION: PEPTIDES FOR ANTIMICROBIAL AND STERILIZING  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bryan Cave LLP  
STREET: 245 Park Avenue  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10167-0034  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50, 800kb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Microsoft Word 7.0.  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/909,433A  
FILING DATE: 03 JUN 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Leo G. Lenna  
REGISTRATION NUMBER: 42,796  
REFERENCE/DOCKET NUMBER: C36599/105967  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 692-1839  
TELEFAX: (212) 692-1900  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 Amino Acids  
TYPE: Amino Acid  
TOPOLOGY: UNKNOWN  
MOLECULE TYPE: Peptide  
US-08-909-433A-1

Query Match 73.2% Score 30; DB 2; Length 20;  
Best Local Similarity 75.0%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
11111111  
DB 3 GTVLGKK 10

RESULT 13  
PCT-US95-10904-51  
Sequence 51, Application PC/TUS9510904  
GENERAL INFORMATION:  
APPLICANT: PAUL, PREM S.  
APPLICANT: MENG, XIANG-JIN  
APPLICANT: HALBUR, PATRICK G.  
APPLICANT: MOROZOV, IGOR  
TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A  
TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV),  
TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE  
TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10904  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/131,625  
FILING DATE: 05-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M.P.  
REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-021-55X CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 174 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-10904-51

Query Match 73.2% Score 30; DB 5; Length 174;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VLGGKK 8  
11111111  
DB 155 VLGGKK 160

RESULT 14  
US-09-120-365-79  
Sequence 79, Application US/09120365  
Patent No. 610314  
GENERAL INFORMATION:  
APPLICANT: Natori, Shunji  
TITLE OF INVENTION: NEW PROTEASE  
FILE REFERENCE: 32290-144749  
CURRENT APPLICATION NUMBER: US/09/120,365  
CURRENT FILING DATE: 1998-07-22  
EARLIER APPLICATION NUMBER: JP 9-333 474  
EARLIER FILING DATE: 1997-11-18  
NUMBER OF SEQ ID NOS: 101  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 79  
LENGTH: 282  
TYPE: PRT  
ORGANISM: Homo sapien cathepsin B  
US-09-120-365-79

Query Match 73.2% Score 30; DB 3; Length 282;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
11111111  
DB 60 GTVLGKK 67

RESULT 15  
US-09-515-039-79  
Sequence 79, Application US/09515039  
Patent No. 621459  
GENERAL INFORMATION:  
APPLICANT: Natori, Shunji  
TITLE OF INVENTION: NEW PROTEASE  
FILE REFERENCE: 32290-144749  
CURRENT APPLICATION NUMBER: US/09/515,039  
CURRENT FILING DATE: 2000-03-06  
EARLIER APPLICATION NUMBER: JP 9-333 474  
EARLIER FILING DATE: 1997-11-18  
NUMBER OF SEQ ID NOS: 101  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 79  
LENGTH: 282  
TYPE: PRT  
ORGANISM: Homo sapien cathepsin B  
US-09-515-039-79

Query Match 73.2%; Score 30; DB 4; Length 282;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTVLCGKK 8  
|||  
Db 60 GTFLGPK 67

Search completed: July 30, 2002, 15:34:52  
Job time: 237 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:35:50 ; Search time 51.53 Seconds

(without alignments)  
14.918 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGSKK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	251	2	E82027
2	35	85.4	339	1	KHMSB
3	34	82.9	130	2	S63533
4	34	82.9	214	2	H82745
5	33	80.5	227	2	B97860
6	33	80.5	244	2	C95846
7	33	80.5	316	2	S63385
8	33	80.5	998	2	S31735
9	32	78.0	150	2	T38477
10	32	78.0	288	2	D64694
11	32	78.0	337	2	A96953
12	32	78.0	375	2	D75568
13	32	78.0	434	2	D66760
14	32	78.0	496	2	A83107
15	32	78.0	717	2	T06041
16	32	78.0	786	2	A35466
17	32	78.0	923	2	A39596
18	32	78.0	923	2	I53280
19	32	78.0	930	2	A25923
20	32	78.0	933	1	QRHUP
21	32	78.0	1127	2	T03105
22	31	75.6	108	2	S69294
23	31	75.6	148	2	S10655
24	31	75.6	157	2	F83749
25	31	75.6	159	2	A96962
26	31	75.6	175	2	S76341
27	31	75.6	178	2	E88637
28	31	75.6	199	2	G75153
29	31	75.6	202	2	A71195

30	31	75.6	220	2	E71857	probable outer mem
31	31	75.6	230	2	C64658	outer membrane pro
32	31	75.6	283	2	G98020	hypothetical prote
33	31	75.6	306	2	AH0385	geranyltransferase
34	31	75.6	339	1	KHRTB	cathepsin B (EC 3.
35	31	75.6	354	2	T44934	moa protein limpo
36	31	75.6	361	2	A75525	conserved hypotet
37	31	75.6	367	2	T16913	hypothetical prote
38	31	75.6	431	2	AH2861	hypothetical prote
39	31	75.6	444	2	S42677	HES1 protein - yea
40	31	75.6	443	2	T17970	hypothetical prote
41	31	75.6	462	2	G97638	penicillin-binding
42	31	75.6	475	2	S73746	MG294 homolog A05-
43	31	75.6	479	2	F87474	major facilitator
44	31	75.6	484	2	G70846	hypothetical glyci
45	31	75.6	495	2	T42758	double-stranded RN

#### ALIGNMENTS

RESULT 1.  
Probable MAPB alternative C-terminus NMA0319 [Imported] - Neisseria meningitidis (str C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: E82027  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; MO Nature 404, 502-506, 2000.  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: A81775; MUID:20222556  
A:Accession: E82027  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-251 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; MID:97378778; PIDN:CA83624.1; PID:9737  
A:Experimental source: serogroup A, strain Z2491  
A:Genetics:  
A:Gene: matB', NMA0319

Query Match 85.4%; Score 35; DB 2; Length 251;  
Best local similarity 85.4%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSKK 8  
||| |||  
DB 68 GTVMGSKK 75

RESULT 2.  
KHMSB  
cathepsin B (EC 3.4.22.1) precursor - mouse  
N:Alternate names: precocathepsin B  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1992 #sequence\_revision 26-Apr-1996 #text\_change 18-Jun-1999  
C:Accession: A8458; A49826; B26498; S12801; PS0360  
R:Qian, F.; Frankel, A.; Chan, S.D.; Steiner, D.F.  
DNA Cell Biol. 10, 159-168, 1991  
A:Title: The structure of the mouse cathepsin B gene and its putative promoter.  
A:Reference number: A38458; MUID:91190267  
A:Accession: A38458  
A:Molecule type: DNA  
A:Residues: 1-339 <QIT>  
A:Cross-references: GB:MG3999  
R:Qian, F.; Frankel, A.; Steiner, D.F.; Bajkowski, A.S.; Chan, S.J.  
Anticancer Res. 11, 1445-1451, 1991  
A:Title: Characterization of multiple cathepsin B mRNAs in murine B16a melanoma.  
A:Reference number: A49826; MUID:92082172  
A:Accession: A49826  
A:Molecule type: mRNA  
A:Residues: 1-339 <QIT>

A:Cross-references: GB:S69034; NID:g239906; PIDN:AAB20536.1; PID:g239907  
 A:Experimental source: B16a melanoma  
 A:Note: Sequence extracted from B16a melanoma (NCIN:69034, NCHP:69039)  
 R:Chan, S.J.; San Segundo, B.; McCormick, M.B.; Steiner, D.F.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 7721-7725, 1986  
 A:Title: Nucleotide and predicted amino acid sequences of cloned human and mouse preproc  
 A:Reference number: A26498; MUID:87017021  
 A:Accession: B26498  
 A:Molecule type: mRNA  
 A:Residues: 1-159, 'N', 161-173, 'D', 175-176, 'I', 178-283, 'V', 285-339 <CHN>  
 A:Cross-references: GB:M14222; NID:g192841; PIDN:AA37494.1; PID:g309202  
 R:Ferrera, M.; Wojcik, F.; Rhaisst, H.; Mordier, S.; Roux, M.P.; Bechet, D.  
 FEBS Lett. 273, 195-199, 1990  
 A:Title: Gene structure of mouse cathepsin B.  
 A:Reference number: S12901; MUID:91032179  
 A:Accession: S12901  
 A:Molecule type: DNA  
 A:Residues: 1-159, 'N', 161-173, 'D', 175-176, 'I', 178-283, 'V', 285-339 <FER>  
 R:Fiemert, C.; Closs, E.I.; Silbermann, M.; Effle, V.; Strauss, P.G.  
 Gene 103, 259-261, 1991  
 A:Title: Isolation of a cathepsin B-encoding cDNA from murine osteogenic cells.  
 A:Reference number: PS0360; MUID:91365255  
 A:Accession: PS0360  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 314-339 <FRI>  
 A:Cross-references: EMBL:X54966; NID:g50596; PIDN:CA38713.1; PID:g50597  
 C:Genetics:  
 A:Note: single copy gene  
 C:Function:  
 A:Description: broad specificity endopeptidase and peptidyl dipeptidase  
 A:Pathway: Intracellular protein degradation  
 C:Superfamily: papain  
 C:Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome; protein degradation  
 F:1-17/Domain: signal sequence #status predicted <Sig>  
 F:80-339/Product: cathepsin B #status predicted <MAT>  
 F:8, 192/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:93-122, 105-150, 141-207, 142-146, 179-211, 187-198/Disulfide bonds: #status predicted  
 F:108, 278/Active site: Cys, His, Asn #status predicted

Query Match 85.4%; Score 35; DB 1; Length 339;  
 Best Local Similarity 87.5%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GTVLGCK 8  
 |||||  
 Db 60 GTVLGCK 67

RESULT 3  
 S63533  
 C:Species: Entamoeba histolytica  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999  
 C:Accession: S63533; S63474  
 R:Binder, M.; Ortner, S.; Erben, H.; Scheiner, O.; Wiedermann, G.; Valenta, R.; Duchene,  
 Eur. J. Biochem. 233, 976-981, 1995  
 A:Title: The basic isoform of profilin in pathogenic Entamoeba histolytica: cDNA cloning  
 A:Reference number: S63474; MUID:96085166  
 A:Accession: S63474  
 A:Molecule type: mRNA  
 A:Residues: 1-130 <BIN>  
 A:Cross-references: EMBL:X90911; NID:q1070154; PID:q1070155  
 A:Accession: S63474  
 A:Molecule type: protein  
 A:Residues: 74-84; 86-94 <BIN>  
 C:Superfamily: profilin  
 C:Keywords: actin binding

Query Match 82.9%; Score 34; DB 2; Length 130;

Best Local Similarity 75.0%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GTVLGCK 8  
 |||||  
 Db 61 GTVLGCK 68

RESULT 4  
 B82745  
 A:Title: Xylella fastidiosa consortium of the Organization for Nucleotide Seq  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82745  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-214 <SIM>  
 A:Cross-references: GB:AE003931; GB:AE003849; NID:g9105834; PIDN:AAF83728.1; GSPDB:CN  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carier  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinca, A.P.; Ferreira, A.J.S.  
 Submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
 chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miraca, E.C.; Miyaki, C.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava  
 M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0918

Query Match 82.9%; Score 34; DB 2; Length 214;  
 Best Local Similarity 85.7%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GTVLGCK 7  
 |||||  
 Db 59 GTVLGCK 65

RESULT 5  
 B97860  
 C:Species: Rickettsia conorii  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
 C:Accession: B97860  
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.;  
 Science 293, 2093-2098, 2001  
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A:Reference number: A97700; MUID:21442074; PMID:11557893  
 A:Accession: B97860  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-227 <KOR>  
 A:Cross-references: GB:AE006914; PIDN:ALU03820.1; PID:g15620420; GSPDB:GN00173  
 C:Genetics:  
 A:Gene: RC1282  
 C:Superfamily: Rickettsia prowazekii hypothetical protein RP828

Query Match 80.5%; Score 33; DB 2; Length 227;

Best Local Similarity 71.4%; Pred. No. 64;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7  
||:||||  
Db 82 GTVLGK 88

RESULT 6

C95846  
hypothetical protein [imported] - *Sinorhizobium meliloti* (strain 1021) megaplasmid pSymB  
C:Species: *Sinorhizobium meliloti*  
C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: C95846  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001.  
A:Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: C95846  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-244 <KUR>  
A:Cross-References: GB:AL591985; PIDN:CAC48435.1; PID:915139907; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Telauere,  
heault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMD20032  
A:Genome: plasmid

Query Match 80.5%; Score 33; DB 2; Length 244;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7  
||:||||  
Db 214 GTVLGK 220

RESULT 7

S63385  
hypothetical protein YNR054c - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein N3491  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 05-Nov-1999  
C:Accession: S63385  
R:Pohl, T.M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63346  
A:Accession: S63385  
A:Molecule type: DNA  
A:Residues: 1-316 <POH>  
A:Cross-References: EMBL:Z71669; NID:g1302571; PID:e239595; PID:g1302572; GSPDB:GN00014;  
C:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YNR054c  
A:Map position: 14R

Query Match 80.5%; Score 33; DB 2; Length 316;  
Best Local Similarity 62.5%; Pred. No. 88;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
|:|||||

Db 189 GNIGGK 196

RESULT 8

S31735  
NAD+ ADP-ribosyltransferase (EC 2.4.2.30) - African clawed frog (fragment)  
N:Alternate names: poly ADP-ribose polymerase  
C:Species: *Xenopus laevis* (African clawed frog)  
C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Jun-2000  
C:Accession: S31735; PNO495  
R:Saulier-Le Drian, B.M.  
submitted to the EMBL Data Library, May 1992  
A:Reference number: S31735  
A:Accession: S31735  
A:Molecule type: mRNA  
A:Residues: 1-998 <SAD>  
A:Cross-References: EMBL:Z12139; NID:964967; PIDN:CAA78126.1; PID:g1334661  
R:Ozawa, Y.; Uchida, K.; Uchida, M.; Aml, Y.; Kushida, S.; Okada, N.; Miwa, M.  
Biochem. Biophys. Res. Commun. 193, 119-125, 1993  
A:Title: Isolation of cDNAs encoding the catalytic domain of poly(ADP-ribose)polymera  
A:Reference number: PNO494; MUID:93277538  
A:Accession: PNO495  
A:Molecule type: mRNA  
A:Residues: 742-745, 'E', 747-876 <OZA>  
C:Comment: This zinc-finger protein plays a role in DNA repair, cell growth, and diff  
C:Superfamily: NAD+ ADP-ribosyltransferase  
C:Keywords: DNA binding; glycosyltransferase; hexosyltransferase; NAD; nucleus; pento

Query Match 80.5%; Score 33; DB 2; Length 998;  
Best Local Similarity 75.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
||:||||  
Db 576 GTVLGK 583

RESULT 9

T38477  
very hypothetical protein SPAC29A4.12c - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38477  
R:Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1997  
A:Reference number: Z21796  
A:Accession: T38477  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-150 <BRG>  
A:Cross-References: EMBL:Z97210; PIDN:CAH10139.1; GSPDB:GN00006; SPDB:SPAC29A4.12c  
C:Experimental source: strain 972h-; cosmid c29A4  
C:Genetics:  
A:Gene: SPDB:SPAC29A4.12c  
A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 150;  
Best Local Similarity 62.5%; Pred. No. 68;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
||:||||  
Db 101 GSIFGK 108

RESULT 10

D64694  
hypothetical protein HP1396 - *Helicobacter pylori* (strain 26695)  
C:Species: *Helicobacter pylori*  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: D64694

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A:Reference number: A64520; MUID:97394467  
 A:Accession: D64694  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-288 <TOM>  
 A:Cross-references: GB:AE000639; GB:AE000511; NID:g2314560; PIDN:AA08441.1; PID:g231457

Query Match 78.0%; Score 32; DB 2; Length 288;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TWLGKK 8  
 DB 219 TLGKK 225

RESULT 11  
 A96953  
 methyl-accepting chemotaxis protein [imported] - *Clostridium acetobutylicum*  
 C:Species: *Clostridium acetobutylicum*  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: A96953  
 R:Noelling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: A96953  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-337 <KOR>  
 A:Cross-references: GB:AE001437; PIDN:AAK78412.1; PID:g15023287; GSPDB:GN00168  
 A:Experimental source: *Clostridium acetobutylicum* ATCC824  
 C:Genetics:  
 A:Gene: CAC0432

Query Match 78.0%; Score 32; DB 2; Length 337;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7  
 DB 109 GTVLGQ 115

RESULT 12  
 D75568  
 N-acetylaminic acid racemase - *Deinococcus radiodurans* (strain R1)  
 C:Species: *Deinococcus radiodurans*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: D75568  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: D75568  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-375 <WHI>  
 A:Cross-references: GB:AE001867; GB:AE000513; NID:g6457693; PIDN:AA09631.1; PID:g645769  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DG0044

A:Map position: 1  
 C:Superfamily: muconate cyclisomerase

Query Match 78.0%; Score 32; DB 2; Length 375;  
 Best Local Similarity 75.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
 DB 127 GTVLGK 134

RESULT 13  
 D96760  
 hypothetical protein T9L24.41 [imported] - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D96760  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Fedorapfel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,ansen, N.F.; Hughes, B.; Hultar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo,ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A6141; MUID:21016719  
 A:Accession: D96760  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-434 <SNO>  
 A:Cross-references: GB:AE005173; NID:g1120791; PIDN:AA030971.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T9L24.41  
 A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 434;  
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8  
 DB 324 GTVLGK 331

RESULT 14  
 AB3107  
 conserved hypothetical protein PAA308 [imported] - *Pseudomonas aeruginosa* (strain PAO  
 C:Species: *Pseudomonas aeruginosa*  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: AB3107  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; L,ory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pa  
 A:Reference number: AB2950; MUID:20437337  
 A:Accession: AB3107  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-496 <SNO>  
 A:Cross-references: GB:AE004847; GB:AE004091; NID:g9950528; PIDN:AA07696.1; GSPDB:GN  
 A:Experimental source: strain PAO1  
 C:Genetics:  
 A:Gene: PAA308

Query Match 78.0%; Score 32; DB 2; Length 496;  
 Best Local Similarity 62.5%; Pred. No. 2.1e+02;



Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 8

Db 487 GSLMGSK 494

RESULT 15

T06041

Hypothetical protein T24A18.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06041

R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Be submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15484

A:Accession: T06041

A:Molecule type: DNA

A:Residues: 1-717 <BEV>

A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.60

A:Experimental source: cultivar Columbia; BAC clone T24A18

C:Genetics:

A:Gene: ATSP:T24A18.60

A:Map position: 4

A:Introns: 7/3; 239/1; 287/3; 398/3

Query Match

Best Local Similarity 78.0%; Score 32; DB 2; Length 717;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSK 7

Db 207 GTVFGSK 213

Search completed: July 30, 2002, 15:35:52  
Job time: 257 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:58 ; Search time 27.4 Seconds

(without alignments)  
11.305 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGSK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	339	1	CATB_MOUSE
2	34	82.9	130	1	PROF_ENTHI
3	33	80.5	316	1	YNSV_YEAST
4	33	80.5	998	1	PPOL_XENLA
5	32	78.0	150	1	YDPC_SCHPO
6	32	78.0	180	1	PRGR_MACRU
7	32	78.0	375	1	RCU1_ARATH
8	32	78.0	377	1	PRGR_SHEEP
9	32	78.0	786	1	PRGR_CHICK
10	32	78.0	923	1	PRGR_MOUSE
11	32	78.0	923	1	PRGR_MOUSE
12	32	78.0	930	1	PRGR_RABIT
13	32	78.0	933	1	PRGR_HUMAN
14	31	75.6	148	1	YORX_PYRMO
15	31	75.6	339	1	CATB_RAT
16	31	75.6	367	1	YOV4_CAEEL
17	31	75.6	372	1	PEX3_MOUSE
18	31	75.6	372	1	PEX3_RAT
19	31	75.6	373	1	PEX3_CRIO
20	31	75.6	373	1	PEX3_HUMAN
21	31	75.6	434	1	HES1_YEAST
22	31	75.6	475	1	Y294_MYCPN
23	31	75.6	521	1	GAG_HYD2
24	31	75.6	559	1	FRB7_WHEAT
25	31	75.6	565	1	DSBD_ECO57
26	31	75.6	565	1	DSBD_ECOLI
27	31	75.6	627	1	GIDA_COXBU
28	31	75.6	628	1	GIDA_BUCAP
29	31	75.6	631	1	GIDA_BUCAP
30	31	75.6	631	1	GIDA_VIBCH
31	31	75.6	635	1	GIDA_SYNT3
32	31	75.6	1144	1	CVA3_HUMAN
33	31	75.6	1144	1	CVA3_RAT

J

34	31	75.6	1914	1	KML5_HUMAN
35	30	73.2	15	1	ASPI_IACSN
36	30	73.2	171	1	VPX_IDV
37	30	73.2	212	1	REP2_RUOLA
38	30	73.2	289	1	AROE_ARCFU
39	30	73.2	299	1	ISPA_ECOLI
40	30	73.2	291	1	YC21_METUA
41	30	73.2	301	1	CUT2_SCHPO
42	30	73.2	325	1	MOCG_RHIME
43	30	73.2	328	1	K6PF_MYCPN
44	30	73.2	339	1	CATB_HUMAN
45	30	73.2	340	1	CATB_CHICK

## ALIGNMENTS

RESULT - 1	ID	CATB_MOUSE	STANDARD:	PRT:	339 AA.
AC	P10605				
DT	01-JUL-1989 (Rel. 11, Created)				
DT	01-FEB-1991 (Rel. 17, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	Cathepsin B precursor (EC 3.4.22.1) (Cathepsin B1).				
CN	CNSB.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91190267; PubMed=2012677;				
RA	Qian F., Franklater A., Chan S.J., Steiner D.F.;				
RT	"The structure of the mouse cathepsin B gene and its putative promoter.";				
RL	DNA Cell Biol. 10:159-168(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91032179; PubMed=2226854;				
RA	Ferrara M., Wojcik F., Rhaissi H., Mordler S., Roux M.-P., Bechet D.;				
RT	"Gene structure of mouse cathepsin B.";				
RL	FEBS Lett. 273:195-199(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=87017021; PubMed=3463966;				
RT	Chan S.J., San Segundo B., McCormick M.B., Steiner D.F.;				
RL	"Nucleotide and predicted amino acid sequences of cloned human and mouse procathesin B cDNAs.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:7721-7725(1986).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92082172; PubMed=1746902;				
RA	Qian F., Franklater A., Steiner D.F., Bajkowski A.S., Chan S.J.;				
RT	"Characterization of multiple cathepsin B mRNAs in murine B16 melanoma.";				
RL	Anticancer Res. 11:1445-1451(1991).				
RN	[5]				
RP	SEQUENCE OF 314-339 FROM N.A.				
RX	MEDLINE=91365255; PubMed=1889751;				
RA	Friemert C., Closs E.I., Silbermann M., Efile V., Strauss P.G.;				
RT	"Isolation of a cathepsin B-encoding cDNA from murine osteogenic cells.";				
RL	Gene 103:259-261(1991).				
CC	-1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.				
CC	-1- CATALYTIC ACTIVITY: HYDROLASES PROTEINS, WITH A SPECIFICITY RESEMBLING THAT OF PAPAIN.				
CC	-1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED BY A DISULFIDE BOND.				
CC	-1- SUBCELLULAR LOCATION: Lysosomal.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE				

Q15746	homo sapien
P82648	lactobacilli
P24123	lactate deh
P13779	kluyveromyc
O27957	archaeoglob
P22939	escherichia
Q58618	methanococc
P21135	schizosacch
P49304	rhizobium m
P75476	mycoplasma
P07858	homo sapien
P43233	gallus galli

## PAPAIN FAMILY OF THIOL PROTEASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M65210; AAA37375.1; -  
 DR EMBL: M65263; AAA37375.1; JOINED.  
 DR EMBL: M65264; AAA37375.1; JOINED.  
 DR EMBL: M65265; AAA37375.1; JOINED.  
 DR EMBL: M65266; AAA37375.1; JOINED.  
 DR EMBL: M65267; AAA37375.1; JOINED.  
 DR EMBL: M65268; AAA37375.1; JOINED.  
 DR EMBL: M65269; AAA37375.1; JOINED.  
 DR EMBL: M4222; AAA37494.1; -  
 DR EMBL: X54966; CAA38713.1; -  
 DR EMBL: S69034; AAB20536.1; -  
 DR PIR: B26498; KMSB.  
 DR PIR: S12901; S12901.  
 DR PIR: A38458; A38458.  
 DR PIR: A49826; A49826.  
 DR HSP: P00787; 1THE.  
 DR MEROPS: C01.060; -  
 DR SWISS-2DPAGE: P10605; MOUSE.  
 DR MGD: MGI:88561; Ct5b.  
 DR InterPro: IPR000668; Peptidase.C1.  
 DR InterPro: IPR00169; ThiolProt\_act\_site.  
 DR Pfam: PF00112; Peptidase.C1; 1.  
 DR PRINTS: PR00705; PAPAIN.  
 DR PROSITE: PS00139; THOL\_PROTEASE\_CYS; 1.  
 DR PROSITE: PS00639; THOL\_PROTEASE\_HIS; 1.  
 DR PROSITE: PS00640; THOL\_PROTEASE\_ASN; 1.  
 DR Hydrolase: Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal.  
 KM SIGNAL 1 17  
 FT PROPEP 18 79 ACTIVATION PEPTIDE.  
 FT CHAIN 80 333 CATHEPSIN B.  
 FT CHAIN 80 126 CATHEPSIN B LIGHT CHAIN (BY SIMILARITY).  
 FT CHAIN 129 333 CATHEPSIN B HEAVY CHAIN (BY SIMILARITY).  
 FT PROPEP 334 339 BY SIMILARITY.  
 FT ACT\_SITE 108 108 BY SIMILARITY.  
 FT ACT\_SITE 278 278 BY SIMILARITY.  
 FT ACT\_SITE 298 298 BY SIMILARITY.  
 FT DISULFID 93 122 BY SIMILARITY.  
 FT DISULFID 105 150 BY SIMILARITY.  
 FT DISULFID 141 207 BY SIMILARITY.  
 FT DISULFID 142 146 BY SIMILARITY.  
 FT DISULFID 179 211 BY SIMILARITY.  
 FT DISULFID 187 198 BY SIMILARITY.  
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 160 160 S -> N (IN REF. 3).  
 FT CONFLICT 174 174 N -> I (IN REF. 3).  
 FT CONFLICT 177 177 V -> I (IN REF. 3).  
 FT CONFLICT 284 284 G -> V (IN REF. 3).  
 SQ SEQUENCE 339 AA; 37280 MW; 9F0A3CDF70A94040 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 339;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGGK 8  
 |||||  
 DB 60 GTVIGGPK 67

RESULT 2  
 PROF\_ENTHI STANDARD; PRT; 130 AA.  
 AC P49230;

DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Profilin.  
 OS Entamoeba histolytica.  
 OC Eukaryota; Entamoebidae; Entamoeba.  
 OX NCBI\_TaxID=5759;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 74-84 AND 86-94.  
 RC STRAIN-SPL-3;  
 RX MEDLINE=96085166; PubMed=8521867;  
 RA Blinder M., Orthen S., Erben H., Scheiner O., Wiedermann G.,  
 RA Valenta R., Duchene M.;  
 RT "The basic isoform of profilin in pathogenic Entamoeba histolytica.  
 RT cDNA cloning, heterologous expression, and actin-binding  
 RT properties";  
 RL Eur. J. Biochem. 233:976-981(1995)

CC -1- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
 CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW  
 CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
 CC -1- SUBUNIT. OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
 CC ACTIN IN A 1:1 RATIO.  
 CC -1- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X90911; CAA62418.1; -  
 DR HSP: P19984; 2ACG.  
 DR InterPro: IPR002097; Profilin.  
 DR Pfam: PF00235; Profilin; 1.  
 DR SMART: SM00392; PROF; 1.  
 DR PROSITE: PS00414; PROFILIN; 1.  
 DR Actin-binding::Cytoskeleton.  
 KM Actin-binding::Cytoskeleton.  
 SQ SEQUENCE 130 AA; 13237 MW; AF50A8E08FA00A6F CRC64;

Query Match 82.9%; Score 34; DB 1; Length 130;  
 Best Local Similarity 75.0%; Pred. No. 8.6;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGGK 8  
 |||||  
 DB 61 GTVIGGK 68

RESULT 3  
 YN8V\_YEAST STANDARD; PRT; 316 AA.  
 ID YN8V\_YEAST  
 AC P53743;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 36.4 kDa protein in POP2-HOL1 intergenic region.  
 GN YN054C OR N3491.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pohl T.M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: 271669; CAA96335.1; -  
 DR SGI: S0005337; YNR054C.1  
 DR InterPro: IPR003955; RRM.2.  
 DR SMART: SM00362; RRM.2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 316 AA; 36408 MW; 5F38B80CDEEC01F CRC64;

Query Match 80.5%; Score 33; DB 1; Length 316;  
 Best Local Similarity 62.5%; Pred. No. 30;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
 I : : : : :  
 Db 189 GNITGKK 196

RESULT 4  
 PROL\_XENLA STANDARD; PRT; 998 AA.  
 AC P31659;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Poly (ADP-ribose) polymerase (EC 2.4.2.30) (PARP) (ADPRT) (NAD(+)-ADP-riboyltransferase) (Poly(ADP-ribose) synthetase) (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RA Saulier-Le Drian B.M.;  
 RL Thesis (1992), University of Rennes, France.  
 RT [2]  
 RX SEQUENCE OF 742-876 FROM N.A.  
 RX MEDLINE=93277538; PubMed=8503897;  
 RA Ozawa Y., Uchida K., Ami Y., Kushida S., Okada N., Miwa M.;  
 RA "Isolation of cDNAs encoding the catalytic domain of poly(ADP-ribose) polymerase from Xenopus laevis and cherry salmon using heterologous oligonucleotide consensus sequences";  
 RT Blochem. Biophys. Res. Commun. 193:119-125(1993).  
 RL -1- FUNCTION: POLY(ADP-RIBOSE) POLYMERASE MODIFIES VARIOUS NUCLEAR PROTEINS BY POLY(ADP-RIBOSYL)ATION. THE MODIFICATION IS DEPENDENT ON DNA AND IS INVOLVED IN THE REGULATION OF VARIOUS IMPORTANT CELLULAR PROCESSES SUCH AS DIFFERENTIATION, PROLIFERATION, AND TUMOR TRANSFORMATION AND ALSO IN THE REGULATION OF THE MOLECULAR EVENTS INVOLVED IN THE RECOVERY OF CELL FROM DNA DAMAGE.  
 CC -1- CATALYTIC ACTIVITY: NAD(+) + [ADP-D-riboyl(N)-acceptor] = nicotinamide + [ADP-D-riboyl(N+1)-acceptor].  
 CC -1- COFACTOR: ZINC, CONTAINS TWO MOLES OF ZINC PER MOLE OF PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OVARY, OOCYTES, AND BRAIN. LOW IN LIVER.  
 CC -1- MISCELLANEOUS: THE ADP-D-RIBOSYL GROUP OF NAD(+) IS TRANSFERRED TO AN ACCEPTOR CARBOXYL GROUP ON A HISTONE OR THE ENZYME ITSELF, AND FURTHER ADP-RIBOSYL GROUPS ARE TRANSFERRED TO THE 2'-POSITION OF THE TERMINAL ADENOSINE MOIETY, BUILDING UP A POLYMER WITH AN AVERAGE CHAIN LENGTH OF 20-30 UNITS.  
 CC -1- SIMILARITY: BELONGS TO THE PARP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BCT DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: 212139; GAA78126.1; -  
 DR EMBL: D13810; BAA02966.1; -  
 DR PIR: S31735; S31735.  
 DR HSSP: P26446; IAN6.  
 DR InterPro: IPR001357; BRC1.  
 DR InterPro: IPR001290; PARP.  
 DR InterPro: IPR004102; PARP-reg.  
 DR InterPro: IPR001510; Znf-PARP.  
 DR Pfam: PF00553; BRC1.  
 DR Pfam: PF00644; PARP.1.  
 DR Pfam: PF02877; PARP-reg.1.  
 DR Pfam: PF00645; zif-PARP.2.  
 DR ProDom: PD004675; Znf-PARP.2.  
 DR SMART: SM00292; BRC1.1.  
 DR PROSITE: PS50172; BRC1.  
 DR PROSITE: PS00347; PARP\_ZN\_FINGER\_1; 2.  
 DR PROSITE: PS50064; PARP\_ZN\_FINGER\_2; 2.  
 KW Transferase; Glycosyltransferase; NAD; DNA-binding; Nuclear protein; ADP-ribosylation; zinc-finger; zinc.  
 FT NON\_TER 1 1  
 FT DNA\_BIND <1 356  
 FT DOMAIN 357 507  
 FT DOMAIN 369 445  
 FT DOMAIN 508 998  
 FT ZN\_FING 8 43  
 FT ZN\_FING 111 148  
 FT DOMAIN 193 195  
 FT DOMAIN 207 212  
 FT DOMAIN 391 391  
 FT MOD\_RES 397 397  
 FT MOD\_RES 419 419  
 FT MOD\_RES 428 428  
 FT MOD\_RES 429 429  
 FT MOD\_RES 445 445  
 FT MOD\_RES 447 447  
 FT MOD\_RES 454 454  
 FT MOD\_RES 467 467  
 FT MOD\_RES 471 471  
 FT MOD\_RES 477 477  
 FT MOD\_RES 495 495  
 FT MOD\_RES 496 496  
 FT MOD\_RES 503 503  
 FT CONFLICT 746 746  
 SQ SEQUENCE 998 AA; 111126 MW; F5A25FAA3366BAE7 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 998;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
 I : : : : :  
 Db 576 GTVLGSKK 583

RESULT 5  
 YDPC\_SCHPO STANDARD; PRT; 150 AA.  
 ID YDPC\_SCHPO  
 AC O14015;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical 16.8 kDa protein C29A4.12c in chromosome I.  
 GN SPAC29A4.12C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: 297210; CAB10139.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 150 AA; 16781 MW; 3CFPCDE23106361 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 150;  
 Best Local Similarity 62.5%; Pred. No. 24;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8  
 I:||||  
 DB 101 GSIFGGRK 108

RESULT 6  
 PRGR\_MACEU  
 ID PRGR\_MACEU STANDARD; PRT; 180 AA.  
 AC P79373;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Progesterone receptor (PR) (Fragment).  
 GN PGR OR NR3C3.  
 OS Macropus eugenii (Tamar wallaby).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
 OX NCBI\_TaxID=9315;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RX MEDLINE=96401260; PubMed=8807636;  
 RA Lim-tio S.S., Keightley M.C., Fletcher T.P., Fuller P.J.;  
 RT "The molecular basis of RU486 resistance in the Tamar Wallaby,  
 RT Macropus eugenii.";  
 RL Mol. Cell. Endocrinol. 119:169-174(1996).  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: S83227; AAB49508.2; -  
 DR HSSP: P06401; 1A2B.  
 DR InterPro: IPR000536; Hormone\_rec\_11g.  
 DR InterPro: IPR001628; zf-C4.

DR Pfam: PF00104; hormone\_rec\_1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; PARTIAL.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Steroid-binding.  
 FT NON\_TER 1 1  
 FT DNA\_BIND <1 16 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING <1 11 C4-TYPE.  
 FT DOMAIN 65 >180 STEROID-BINDING.  
 FT NON\_TER 180 180  
 SQ SEQUENCE 180 AA; 20480 MW; D1618A89061C7BF3 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 180;  
 Best Local Similarity 75.0%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8  
 I:||||  
 DB 15 GMVLGGRK 22

RESULT 7  
 RCL1\_ARATH  
 ID RCL1\_ARATH STANDARD; PRT; 375 AA.  
 AC O9C578;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable RNA 3'-terminal phosphate cyclase-like protein.  
 GN A15622100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=cv. COLUMBIA;  
 RX MEDLINE=21016721; PubMed=11130714;  
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Naito K., Okumura S., Shipo S., Takeuchi C., Wada T.,  
 RA Matanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Habermann K., Murray J., Johnson D., Kohling T., Nelson J.,  
 RA Steneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,  
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N., J.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.,  
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Voiclaert G., Wambolt R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,  
 RA Langham K.-A., McCullagh B., Robben J., Gymnopoulos B., Zimmermann W.,  
 RA Ransperger U., Weider H., Balke K., Wedler E., Peters S.,  
 RA van Staveren M., Dirkse W., Woolfman P., Klein Lankhorst R.,  
 RA Weltzienegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,  
 RA Feldpusch M., Lamberth S., Villarroel R., Giesen J., Ardiles W.,  
 RA Betts O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.F.;  
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:823-826(2000).  
 CC -1- FUNCTION: DOES NOT HAVE CYCLASE ACTIVITY. PLAYS A ROLE IN 40S-  
 CC RIBOSOMAL-SUBUNIT BIOGENESIS IN THE EARLY PRE-RNA PROCESSING  
 CC STEPS AT SITES A0, A1 AND A2 THAT ARE REQUIRED FOR PROPER  
 CC MATURATION OF THE 18S RNA (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RNA 3'-TERMINAL CYCLASE FAMILY.

CC SUBFAMILY 2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL589883; CAC34503.1; -  
 DR InterPro: IPR000228; RTC.  
 DR Pfam: PF01137; RTC; 1.  
 DR Nuclear protein.  
 KW SEQUENCE 375 AA; 40874 MW; 3C7D286D4386D5AF CRC64;  
 SO  
 Query Match 78.0%; Score 32; DB 1; Length 375;  
 Best Local Similarity 71.4%; Pred. No. 55;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTVLGK 7  
 ID 75 GTVLGK 81  
 DB  
 RESULT 8  
 PRGR\_SHEEP STANDARD; PRT; 377 AA.  
 AC Q28590;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Progesterone receptor (PR) (Fragment).  
 GN PGR OR NR3C3.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; ovis.  
 NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ROMANOV; TISSUE-Uterus;  
 RA Madiou T., Tiffoche C., le Gal F., Pelletier J., Thieulant M.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z66555; CAA91447.1; -  
 DR HSSP; P06401; 1A28.  
 DR InterPro: IPR000536; Hormone\_rec\_1lg.  
 DR InterPro: IPR000128; Progesterone\_receptor.  
 DR InterPro: IPR001628; zf-C4.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF02161; Prog\_receptor; 1.  
 DR Pfam; PF00145; zf-C4; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; znF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.

KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Steroid-binding.  
 FT NON\_TER 1  
 FT DOMAIN <1 15  
 FT ZN\_BIND 18 83 MODULATING, PRO-RICH.  
 FT ZN\_FING 18 38 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 54 78 C4-TYPE.  
 FT DOMAIN 128 >377 C4-TYPE.  
 FT NON\_TER 377  
 FT STERIOD-BINDING.  
 KW SEQUENCE 377 AA; 42904 MW; 3141B6587F7493C CRC64;  
 SO  
 Query Match 78.0%; Score 32; DB 1; Length 377;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTVLGK 8  
 ID 82 GNVTLGK 89  
 DB  
 RESULT 9  
 PRGR\_CHICK STANDARD; PRT; 786 AA.  
 AC P07812; Q90946;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Progesterone receptor (PR).  
 GN PGR OR NR3C3.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88166640; PubMed=3443098;  
 RA Gronemeyer H., Turcotte B., Quirin-Sticker C., Bocquel M.T.,  
 RA Meyer M.E., Krozowski Z., Jeltsch J.M., Lerouge T., Garnier J.M.,  
 RA Chambon P.;  
 RT "The chicken progesterone receptor: sequence, expression and  
 RT functional analysis.".  
 RL EMO J. 6:3985-3994(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91042592; PubMed=3153474;  
 RA Conneely O.M., Dobson A.D.W., Tsai M.-J., Beattie W.G., Toft D.O.,  
 RA Hückaby C.S., Zarucki T., Schrader W.T., O'Malley B.W.;  
 RT "Sequence and expression of a functional chicken progesterone  
 RT receptor.".  
 RL Mol. Endocrinol. 1:517-525(1987).  
 RN [3]  
 RP SEQUENCE OF 128-164 FROM N.A.  
 RX MEDLINE=86289413; PubMed=2426779;  
 RA Conneely O.M., Sullivan W.P., Toft D.O., Birnhauser M., Cook R.G.,  
 RA Maxwell B.L., Zarucki T., Greene G.L., Schrader W.T.,  
 RA O'Malley B.W.;  
 RT "Molecular cloning of the chicken progesterone receptor.".  
 RL Science 233:767-770(1986).  
 RN [4]  
 RP SEQUENCE OF 417-490 FROM N.A.  
 RX MEDLINE=86287271; PubMed=2426697;  
 RA Jeltsch J.M., Krozowski Z., Quirin-Sticker C., Gronemeyer H.,  
 RA Simpson R.J., Garnier J.M., Krust A., Jacob F., Chambon P.;  
 RT "Cloning of the chicken progesterone receptor.".  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5424-5428(1986).  
 RN [5]  
 RP DIFFERENCE BETWEEN FORM 1 AND FORM 2.  
 RX MEDLINE=89340509; PubMed=2760059;  
 RA Conneely O.M., Kettelberger D.M., Tsai M.-J., Schrader W.T.,  
 RA O'Malley B.W.;  
 RT "The chicken progesterone receptor A and B isoforms are products of

an alternate translation initiation event.";  
 J. Biol. Chem. 264:14062-14064(1989).  
 [6]  
 MEDLINE-90154085; PubMed-2303486;  
 J. M., Turcotte B., Garnier J.-M., Lerouge T., Krozowski L.,  
 Gronemeyer H., Chambon P.;  
 "Characterization of multiple mRNAs originating from the chicken  
 progesterone receptor gene. Evidence for a specific transcript  
 encoding form A".  
 J. Biol. Chem. 265:3967-3974(1990).  
 -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 -1- SUBCELLULAR LOCATION: Nuclear.  
 -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: A (SHOWN HERE), A', B AND B';  
 ARE PRODUCED BY ALTERNATIVE SPLICING.  
 -1- TISSUE SPECIFICITY: OVIDUCT.  
 -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 NR3 SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: Y00092; CAA68282.1; -  
 EMBL: M13972; AAA49034.1; -  
 EMBL: M37518; AAA49013.1; -  
 EMBL: M37518; AAA49014.1; -  
 EMBL: M14278; AAA49035.1; -  
 EMBL: M14279; AAA49038.1; -  
 EMBL: M14280; AAA49039.1; -  
 EMBL: M32732; AAA49011.1; -  
 EMBL: M31104; AAA49011.1; JOINED.  
 EMBL: M32726; AAA49011.1; JOINED.  
 EMBL: M32727; AAA49011.1; JOINED.  
 EMBL: M32728; AAA49011.1; JOINED.  
 EMBL: M32729; AAA49012.1; JOINED.  
 EMBL: M32730; AAA49012.1; JOINED.  
 EMBL: M32731; AAA49012.1; JOINED.  
 EMBL: M32732; AAA49012.1; JOINED.  
 EMBL: M31104; AAA49012.1; JOINED.  
 EMBL: M32726; AAA49012.1; JOINED.  
 EMBL: M32727; AAA49012.1; JOINED.  
 EMBL: M32728; AAA49012.1; JOINED.  
 EMBL: M32729; AAA49012.1; JOINED.  
 EMBL: M32730; AAA49012.1; JOINED.  
 EMBL: M31104; AAA49009.1; -  
 EMBL: M31104; AAA49010.1; -  
 PIR: A24312; A24312.  
 PIR: A24661; A24661.  
 PIR: S06284; S06284.  
 HSP: P06401; A248.  
 TRANSFAC: T00698; -  
 InterPro: IPR000536; Hormone\_rec\_119.  
 InterPro: IPR000128; Progesterone\_receptor.  
 InterPro: IPR001628; zf-C4.  
 Pfam: PF00104; hormone\_rec\_1.  
 Pfam: PF02161; Prog\_receptor\_1.  
 Pfam: PF00105; zf-C4\_1.  
 PRINTS: PR00047; STEROIDINGER.  
 SMART: SM00430; HOL1; 1.  
 SMART: SM00399; znf\_C4\_1.  
 PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 Receptor: Transcription regulation; DNA-binding; Nuclear protein;  
 zinc-finger; Steroid-binding; Alternative splicing.  
 FT: 1 420 MODULATING\_PRO-RICH.  
 FT: DNA\_BIND 421 486 NUCLEAR\_RECEPTOR-TYPE.

FT ZN\_FING 421 441 C4-TYPE.  
 FT ZN\_FING 457 481 C4-TYPE.  
 FT DOMAIN 487 786 STEROID-BINDING.  
 FT DOMAIN 48 80 ASP/GLU-RICH (ACIDIC).  
 FT VASPLIC 1 127 MISSING (IN ISOFORM B AND ISOFORM B').  
 FT VASPLIC 452 458 OHNYLCA -> TISYCS (IN ISOFORM A AND  
 ISOFORM B').  
 FT VASPLIC 459 786 MISSING (IN ISOFORM A' AND ISOFORM B').  
 FT CONFLICT 58 58 E -> D (IN REF. 2).  
 FT CONFLICT 480 480 K -> N (IN REF. 2).  
 FT CONFLICT 489 489 G -> A (IN REF. 2).  
 FT CONFLICT 577 577 R -> T (IN REF. 2).  
 FT CONFLICT 642 642 M -> I (IN REF. 2).  
 SO SEQUENCE 786 AA; 85743 MW; 65955950BC45ED9 CRC64;  
 Query Match 78.0%; Score 32; DB 1; Length 786;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches: 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CTIVGGRK 8  
 DB 485 GMVLGGRK 492  
 RESULT 10  
 PRGR\_MOUSE STANDARD; PRT; 923 AA.  
 ID PRGR\_MOUSE  
 AC 000175;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Progesterone receptor (PR).  
 GN PRG OR NR3C3 OR PR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RX MEDLINE-9129759; PubMed-2069958;  
 RA Scholt D.R., Shyamala G., Schneider W., Parry G.;  
 RT "Molecular cloning, sequence analyses, and expression of  
 complementary DNA encoding murine progesterone receptor.";  
 RL Biochemistry 30:7014-7020(1991).  
 RP SEQUENCE OF 1-9 FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE-95100931; PubMed-7802637;  
 RA Hagiwara K., Wu-Peng X.S., Funabashi T., Kato J., Pfaff D.W.;  
 RT "Nucleic acid sequence and Dnase hypersensitive sites of the 5'  
 RT region of the mouse progesterone receptor gene".  
 RL Biochem. Biophys. Res. Commun. 205:1093-1101(1994).  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC -1- A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: M68915; AAA39971.1; -  
 EMBL: U12644; AAA6067.1; -  
 PIR: A39596; A39596.



DR HSP; P06401; 1A28.  
 DR MGD; MGI:97567; Paf.  
 DR InterPro; IPR000536; Hormone\_rec\_11g.  
 DR InterPro; IPR000128; Progesterone\_receptor.  
 DR Pfam; PF00104; hormone\_rec.1.  
 DR Pfam; PF02161; Progesterone\_receptor.1.  
 DR Pfam; PF00105; zfc-C4; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR PRINTS; PR00047; STEROIDFINGER.  
 DR SMART; SM00399; znf-C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Steroid-binding; MODULATING, PRO-RICH.  
 FT DOMAIN 1 556  
 FT DNA\_BIND 557 622 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 557 577 C4-TYPE.  
 FT ZN\_FING 593 617 C4-TYPE.  
 FT DOMAIN 671 923 STEROID-BINDING.  
 FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 SQ SEQUENCE 923 AA; 99073 MW; 9415F1ED343BEEF CRC64;

Query Match 78.0%; Score 32; DB 1; Length 923;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
 Db 621 GMVLGKK 628

RESULT 11  
 ID PRGR\_RAT STANDARD; PRT; 923 AA.  
 AC 064449;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Progesterone receptor (PR).  
 GN PGR OR NR3C3.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Placenta;  
 RX MEDLINE=94130817; PubMed=8299566;  
 RA Park-Sarge O.K., Mayo K.E.;  
 RT "Regulation of the progesterone receptor gene by gonadotropins and  
 cyclic adenosine 3',5'-monophosphate in rat granulosa cells.";  
 RL Endocrinology 134:709-718(1994).  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; L16922; AAA19916.1; -;  
 DR HSP; P06401; 1A28.  
 DR InterPro; IPR000536; Hormone\_rec\_11g.

DR InterPro; IPR000128; Progesterone\_receptor.  
 DR InterPro; IPR001628; zfc-C4.  
 DR Pfam; PF00104; hormone\_rec.1.  
 DR Pfam; PF02161; Progesterone\_receptor.1.  
 DR Pfam; PF00105; zfc-C4; 1.  
 DR PRINTS; PR00047; STEROIDFINGER.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; znf-C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Steroid-binding; MODULATING, PRO-RICH.  
 FT DOMAIN 1 556  
 FT DNA\_BIND 557 622 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 557 577 C4-TYPE.  
 FT ZN\_FING 593 617 C4-TYPE.  
 FT DOMAIN 671 923 STEROID-BINDING.  
 FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 SQ SEQUENCE 923 AA; 99407 MW; 05384B9656BF22DC CRC64;

Query Match 78.0%; Score 32; DB 1; Length 923;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8  
 Db 621 GMVLGKK 628

RESULT 12  
 ID PRGR\_RABIT STANDARD; PRT; 930 AA.  
 AC P06186;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Progesterone receptor (PR).  
 GN PGR OR NR3C3.  
 OS Oryctolagus cuniculus (Rabbit).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87067449; PubMed=3538016;  
 RA Logsteil H., Atger M., Mistrant M., Guichon-Mantel A., Meriel C.,  
 RA Logeat F., Benarous R., Milgrom E.;  
 RT "Cloning and sequence analysis of rabbit progesterone-receptor  
 complementary DNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9045-9049(1986).  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M14547; AAA14443.1; -;  
 DR PIR; A25923; A25923.  
 DR HSP; P06401; 1A28.  
 DR TRANSFAC; T00697; -;  
 DR InterPro; IPR000536; Hormone\_rec\_11g.  
 DR InterPro; IPR000128; Progesterone\_receptor.

DR InterPro; IPR001628; zf-C4.  
 DR Pfam; PF00104; hormone\_rec. 1.  
 DR Pfam; PF02161; Prog\_receptor; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00047; STEROIDINER.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; ZNF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 zinc-finger; Steroid-binding.  
 FT DOMAIN 1 565 MODULATING, PRO-RICH.  
 FT DNA\_BIND 568 633 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 568 588 C4-TYPE.  
 FT ZN\_FING 604 628 C4-TYPE.  
 FT DOMAIN 678 930 STEROID-BINDING.  
 FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 SO SEQUENCE 930 AA; 98666 MW; 644FFAC13BF2F883 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 930;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GYVLGGRK 8  
 DB 632 GMYLGGGRK 639

RESULT 13  
 PRGR\_HUMAN STANDARD; PRT; 933 AA.  
 AC P06401; Q90P7; 1  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Progesterone receptor (PR).  
 GN PGR OR NR3C3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE-90228361; PubMed-2328727;  
 RA Kastner P., Krust A., Turcotte B., Stropp U., Tora L., Gronemeyer H.,  
 Chambon P.;  
 RT "Two distinct estrogen-regulated promoters generate transcripts  
 encoding the two functionally different human progesterone receptor  
 forms A and B.";  
 RT EMBO J. 9:1603-1614(1990).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE-87184565; PubMed-3551956;  
 RA Mirshahi M., Atger M., D'Auriol L., Loosfelt H., Meriel C.,  
 Fridlansky F., Guichon-Mantel A., Galibert F., Milgrom E.;  
 RT "Complete amino acid sequence of the human progesterone receptor  
 deduced from cloned cDNA.";  
 RT Biochem. Biophys. Res. Commun. 143:740-748(1987).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RA Kieback D.G., Agoulnik I.U., Tong X.-W.;  
 RP Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 682-933.  
 RX MEDLINE-98282128; PubMed-9620806;  
 RA Williams S.P., Sigler P.B.;  
 RT "Atomic structure of progesterone complexed with its receptor.";  
 RT Nature 393:392-396(1998).  
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use, by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X51730; CA36018.1; -  
 DR EMBL; M15716; AAA60081.1; -  
 DR EMBL; AF016381; AA01587.1; -  
 DR PIR; A03245; Q8RHP.  
 DR PIR; S09971; S09971.  
 DR PDB; 1A28; 15-JUL-98.  
 DR TRASNPA; T00696; -  
 DR MIM; 264080; -  
 DR InterPro; IPR000536; Hormone\_rec\_1lg.  
 DR InterPro; IPR000128; Progesterone\_receptor.  
 DR InterPro; IPR001723; Steroidhormone\_receptor.  
 DR InterPro; IPR001628; zf-C4.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF02161; Prog\_receptor; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00398; STEROIDINER.  
 DR PRINTS; PR00047; STEROIDINER.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; ZNF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW zinc-finger; Steroid-binding; Phosphorylation; Alternative splicing;  
 KW 3D-structure.  
 FT DOMAIN 1 566 MODULATING, PRO-RICH.  
 FT DNA\_BIND 567 632 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 567 587 C4-TYPE.  
 FT ZN\_FING 603 627 C4-TYPE.  
 FT DOMAIN 681 933 STEROID-BINDING.  
 FT DOMAIN 183 187 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 41 41 PHOSPHORYLATION (BY CK).  
 FT MOD\_RES 227 227 PHOSPHORYLATION (BY CK).  
 FT MOD\_RES 232 232 PHOSPHORYLATION (BY CK).  
 FT MOD\_RES 552 552 PHOSPHORYLATION (BY CK).  
 FT MOD\_RES 793 793 PHOSPHORYLATION (BY CK).  
 FT VARSPIC 1 164 MISSING (IN ISOFORM B).  
 FT CONFLICT 226 226 G -> S (IN REF. 1).  
 FT CONFLICT 256 256 V -> S (IN REF. 1).  
 FT CONFLICT 344 344 S -> T (IN REF. 1).  
 FT CONFLICT 660 660 L -> V (IN REF. 2).  
 SO SEQUENCE 933 AA; 98994 MW; 80414B7F1317F8E CRC64;

Query Match 78.0%; Score 32; DB 1; Length 933;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GYVLGGRK 8  
 DB 631 GMYLGGGRK 638

RESULT 14  
 ID YOKX\_PYRMO STANDARD; PRT; 148 AA.  
 AC P20298;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein in GAPDH 3' region (ORF X) (Fragment).

OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN NCB1\_TaxID=2262;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 3773;  
 RX MEDLINE=90330536; PubMed=2165475;  
 RA Zwickl P., Fabry S., Bodegahn C., Haas A., Hensel R.;  
 RT "Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic  
 archaeobacterium Pyrococcus woesei: characterization of the enzyme,  
 RT cloning and sequencing of the gene, and expression in Escherichia  
 coli";  
 RL J. Bacteriol. 172:4329-4338(1990).  
 DR PIR: S10655; S10655.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 148 AA; 15256 MW; 8C7E51A072DF5D50 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTVLCG 6  
 |||||  
 Db 89 GTVLCG 94

RESULT 15  
 CATB\_RAT STANDARD; PRT; 339 AA.  
 AC P00787;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Cathepsin B precursor (EC 3.4.22.1) (Cathepsin B1) (RS6-2).  
 CS CTAB.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Mammary gland;  
 RX MEDLINE=95094788; PubMed=8001549;  
 RA Guenette R.S., Woolbroek M., Wong K., Wong P., Tenniwood M.;  
 RT "Cathepsin B, a cysteine protease implicated in metastatic  
 RT progression, is also expressed during regression of the rat prostate  
 RT and mammary glands";  
 RL Eur. J. Biochem. 226:311-321(1994).  
 RN [2]  
 RP SEQUENCE OF 69-339 FROM N.A.  
 RX MEDLINE=85190489; PubMed=2986112;  
 RA San Segundo B., Chan S.J., Steiner D.F.;  
 RT "Identification of cDNA clones encoding a precursor of rat liver  
 RT cathepsin B";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2320-2324(1985).  
 RN [3]  
 RP SEQUENCE OF 80-126 AND 129-333.  
 RC TISSUE-Liver;  
 RX MEDLINE=83221657; PubMed=6574504;  
 RA Takio K., Towatari T., Katunuma N., Teller D.C., Titani K.;  
 RT "Homology of amino acid sequences of rat liver cathepsins B and H  
 RT with that of papain";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3666-3670(1983).  
 RN [4]  
 RP PROCESSING.  
 RX MEDLINE=92348471; PubMed=1639824;  
 RA Rowan A.D., Mason P., Mach L., Mort J.S.;  
 RT "Rat procathepsin B. Proteolytic processing to the mature form in  
 RT vitro";  
 RL J. Biol. Chem. 267:15993-15999(1992).  
 RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=95197558; PubMed=7890671;  
 RA Uta Z., Hasnain S., Hirama T., Lee X., Mort J.S., To R., Huber C.P.;  
 RT "Crystal structures of recombinant rat cathepsin B and a cathepsin B-  
 RT inhibitor complex. Implications for structure-based inhibitor  
 RT design";  
 RL J. Biol. Chem. 270:5527-5533(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 18-339.  
 RX MEDLINE=96311276; PubMed=8740363;  
 RA Cygler M., Sivaraman J., Grochulski P., Coulombe R., Storer A.C.,  
 RA Mort J.S.;  
 RT "Structure of rat procathepsin B: model for inhibition of cysteine  
 RT protease activity by the proregion";  
 RL Structure 4:405-416(1996).  
 CC -1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN  
 CC INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO  
 CC BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES PROTEINS, WITH A SPECIFICITY  
 CC RESEMBLING THAT OF PAPAIN.  
 CC -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED  
 CC BY A DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
 CC PAPAIN FAMILY OF THIOL PROTEASES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X82396; CAA57792.1; -  
 DR EMBL: M1305; AAA40993.1; -  
 DR PIR: A00977; KHRTB.  
 DR PDB: 1CPT; 07-DEC-95.  
 DR PDB: 1MR; 11-JAN-97.  
 DR PDB: 1CTE; 31-JUL-95.  
 DR PDB: 1THE; 10-MAR-96.  
 DR MEROPS: C01.060; -  
 DR Interpro: IPR000668; Peptidase\_C1.  
 DR Interpro: IPR000169; Thiolprot\_act\_site.  
 DR Pfam: PF00112; Peptidase\_C1; 1.  
 DR PRINTS: PR00705; PAPAIN.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
 DR PROSITE: PS00640; THIOL\_PROTEASE\_ASN; 1.  
 DR KEGG: K01060; -  
 KW Hydrolyase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal;  
 KW 3D-structure.  
 FT SIGNAL 1..17  
 FT PROPEP 18..79  
 FT CHAIN 80..333  
 FT CHAIN 80..126  
 FT CHAIN 129..333  
 FT PROPEP 334..339  
 FT ACT\_SITE 108..108  
 FT ACT\_SITE 278..278  
 FT ACT\_SITE 298..298  
 FT DISULFID 93..122  
 FT DISULFID 105..150  
 FT DISULFID 141..207  
 FT DISULFID 142..146  
 FT DISULFID 179..211  
 FT DISULFID 187..198  
 FT CARBOHYD 192..192  
 FT VARIAT 302..302  
 FT CONFLICT 159..159  
 FT SEQUENCE 339 AA; 37470 MW; 925E2E38C2B03CDA CRC64;

Query Match 75.6%; Score 31; DB 1; Length 339;

Best Local Similarity 100.0%; Pred. NO. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVUGG 6  
|||||  
Db 60 GTVUGG 65

Search completed: July 30, 2002, 15:37:59  
Job time: 313 sec

No.	Score	Match Length	DB	ID	Description
1	37	90.2	525	09C214	09C214 neurospora
2	35	85.4	251	09JWL1	09JWL1 neisseria m
3	35	85.4	286	01 O9DCV0	09dcv0 musc musculus
4	35	85.4	509	02 O9X6P3	09x6p3 neisseria g
5	34	82.9	214	09PEW0	09pew0 xylella fas
6	33	80.5	193	10 O9LJG6	09ljg6 arabidopsis
7	33	80.5	227	10 O9ZG42	09zga2 rickettsia
8	33	80.5	244	16 O9ZXC7	09zxc7 rhizobium m
9	33	80.5	537	09Y154	09y154 trophoblast
10	33	80.5	545	09VA35	09va35 trophoblast
11	33	80.5	607	13 O9PS81	09psl xenopus. na
12	33	80.5	689	10 O9C9J3	09c9j3 arabidopsis
13	32	78.0	64	10 O9PP75	09pp75 oryza sativ
14	32	78.0	153	11 O9Z3G6	09z3g6 mesocricetus
15	32	78.0	174	12 O9WMH2	09wmh2 porcine rep
16	32	78.0	174	12 O9PZ70	09pzt70 porcine rep

17	32	78.0	179	16	09A114	09A114 streptococ
18	32	78.0	254	2	09Z445	09Z445 pseudomonas
19	32	78.0	288	16	02S546	02S546 helicobacte
20	32	78.0	298	6	02R547	02R547 ovis aries
21	32	78.0	337	16	097LX2	097LX2 clostridium
22	32	78.0	349	16	09ZNV6	09ZNV6 rhizobium m
23	32	78.0	360	13	04Z274	04Z274 crocodylus
24	32	78.0	375	16	09RTA6	09RTA6 delnoccocus
25	32	78.0	376	4	09H119	09H119 homo sapien
26	32	78.0	434	10	09FX33	09FX33 aradiposistis
27	32	78.0	496	16	09HW92	09HW92 pseudomonas
28	32	78.0	583	13	09DBV4	09DBV4 xenopus lae
29	32	78.0	717	10	09T045	09T045 arabidopsis
30	32	78.0	732	13	09DDU9	09DDU9 xenopus lae
31	32	78.0	939	6	09GLM0	09GLM0 canis famil
32	32	78.0	960	5	09VTK6	09VTK6 drosophila
33	32	78.0	1078	10	09AFR5	09AFR5 zea mays (m
34	32	78.0	1127	12	036360	036360 alcelaphine
35	32	78.0	10917	2	09JNM6	09JNM6 streptomyce
36	32	78.0	11096	2	09LW43	09LW43 streptomyce
37	31	75.6	77	16	09R8T3	09R8T3 rhizobium l
38	31	75.6	9	15	09E117	09E117 human immun
39	31	75.6	108	3	013531	013531 saccharomyc
40	31	75.6	110	11	09JK77	09JK77 mus musculus
41	31	75.6	140	11	09EBC6	09EBC6 rattus norv
42	31	75.6	149	11	09JHU7	09JHU7 rattus norv
43	31	75.6	157	16	09KEQ1	09KEQ1 bacillus ha
44	31	75.6	159	16	097LQ0	097LQ0 clostridium
45	31	75.6	175	16	055671	055671 synecocysts

## ALIGNMENTS

RESULT	1
ID	09C214
AC	09C214;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	RELATED TO DNA POLYMERASE DELTA SMALL SUBUNIT.
GN	93G11.080.
OS	Neurospora crassa.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC	Sordariales; Sordariaceae; Neurospora.
OX	NCBI_TaxID=5141;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Schulte U., Allyn V., Hohnsels J., Brandt P., Fartmann B., Holland R.
RA	Wakatuna G., Mewes H.W., Mannhaupt G.;
RL	submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RA	German Neurospora genome project;
RL	submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AL513443; CAC28660.2; -
DR	SEQUENCE 525 AA; 58146 MW; 7FF70EEFD05C66F04 CRC64;

Query Match	90.2%	Score 37	DB 3	Length 525
Best local Similarity	87.5%	Pred. No. 48		
Matches 7; Conservative	1	Mismatches 0	Indels 0	Gaps 0

OY	1	GTVLGSK	8		
DB	74	GTVLGSK	81		
RESULT 2					
09JUL1					
09JUL1					
AC					
09JUL1					
PRELIMINARY;					
PRT;					
251 AA					

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE PUTATIVE MAFB ALTERNATIVE C-TERMINUS (FRAGMENT).  
 GN MAFB OR NMA0319.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
 RA Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,  
 RA Rajadaram M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491."  
 RL Nature 404:502-506(2000).  
 DR EMBL: AL162752; CAB83624.1; -.  
 KW Complete proteome.  
 FT NON\_TER  
 SQ SEQUENCE 251 AA; 27783 MW; E6E47DF3CEDF2C7C CRC64;  
 PT  
 S0  
 Query Match 85.4%; Score 35; DB 16; Length 251;  
 Best Local Similarity 87.5%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GTVLGGRK 8  
 Db 68 GTVWGGRK 75  
 RESULT 3  
 ID Q9DCV0 PRELIMINARY; PRT; 286 AA.  
 AC Q9DCV0;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,  
 DE CLONE:0610010D10, FULL INSERT SEQUENCE.  
 GN CTSS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
 RA Kuenl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata K., Storch K.F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 FT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).  
 DR EMBL: AK002456; BAB22113.1; -.  
 DR HSSP: P00787; ITHE.  
 DR MGD: MG1:88561; CtSD.  
 DR InterPro: IPR000668; Peptidase\_C1.  
 DR InterPro: IPR000169; Thiolprot\_act\_site.  
 DR Pfam: PF00112; Peptidase\_C1; 1.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
 KW Hydrolyase; Thiol protease.  
 SQ SEQUENCE 286 AA; 31091 MW; CB9F045097126011 CRC64;  
 PT  
 S0  
 Query Match 85.4%; Score 35; DB 11; Length 286;  
 Best Local Similarity 87.5%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GTVLGGRK 8  
 Db 60 GTVLGGRK 67  
 RESULT 4  
 ID Q9X6P3 PRELIMINARY; PRT; 509 AA.  
 AC Q9X6P3;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE ADHESIN MAFB.  
 GN MAFB.  
 OS Neisseria gonorrhoeae.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=485;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MS11;  
 RA Eickernjaeger S., Meyer T.F., Fischer E., Maier J., Manning P.A.,  
 RA Rudel T., Schuenpflug I., Schulz E., Schwan E.T.;  
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF142582; AAD31039.1; -.  
 SQ SEQUENCE 509 AA; 55681 MW; 88B420CEFD8269D4 CRC64;  
 PT  
 S0  
 Query Match 85.4%; Score 35; DB 2; Length 509;  
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GTVLGGRK 8  
 Db 326 GTVWGGRK 333  
 RESULT 5  
 ID Q9PEW0 PRELIMINARY; PRT; 214 AA.  
 AC Q9PEW0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE ACYL-[ACP]-UDP-N-ACETYLGLUCOSAMINE.  
 GN XPC918.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OX Xylella.  
 RN NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinhard F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Perro J.A.,  
 RA Fraga J.S., Franco S.C., Franco M.C., Frome M., Furian L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H., Gomes S.L., Gruber A.,  
 RA Ho P.L., Holsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurume E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.H.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukawa A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Silvestri W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE003931; AAF83728.1; -;  
 DR InterPro: IPR001451; Hexapep\_transf.  
 DR Pfam: PF00132; hexapep; 9.  
 DR PROSITE: PS00101; HEXAPEP\_TRANSFERASES; UNKNOWN\_3.  
 KW Complete proteome.  
 SQ SEQUENCE 214 AA; 22275 MW; 57AE135939849397 CRC64;

Query Match 82.9%; Score 34; DB 16; Length 214;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GTVLGK 7  
 ||:||||  
 Db 59 GTVLGK 65

RESULT 6  
 ID 09LJG6 PRELIMINARY; PRT; 193 AA;  
 AC 09LJG6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE GENOMIC DNA, CHROMOSOME 3, pl CLONE: MAG2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroids 11; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLMBIA;  
 RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,  
 RL Submitted (OC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLMBIA;  
 RX MEDLINE=20363099; PubMed=10907653;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,  
 RT TAC and BAC clones.";  
 RL DNA Res. 7:217-221(2000).  
 DR EMBL: AP000600; BAB02986.1; -;  
 SQ SEQUENCE 193 AA; 21732 MW; DC8718A166F75161 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 193;

Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTVLGK 8  
 ||:||||  
 Db 34 GTVLGK 41

RESULT 7  
 ID 092G42 PRELIMINARY; PRT; 227 AA.  
 AC 092G42;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE HYPOTHETICAL PROTEIN RC1282.  
 GN RC1282.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OC NCBI\_TaxID=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MALISH 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii,"  
 RL Science 293:2093-2098(2001).  
 DR EMBL: AE008675; AAL03820.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 227 AA; 24449 MW; 315196E8D5E2992 CRC64;

Query Match 80.5%; Score 33; DB 16; Length 227;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTVLGK 7  
 ||:||||  
 Db 82 GTVLGK 88

RESULT 8  
 ID 092XC7 PRELIMINARY; PRT; 244 AA.  
 AC 092XC7;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE HYPOTHETICAL PROTEIN SM20032.  
 GN SM20032.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid symb (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OC NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finn T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Guzy J.,  
 RA Golding B., Pehler A.;  
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL: AL603642; CAC48435.1; -;  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 244 AA; 26022 MW; 848748E827BA876A CRC64;

Query Match 80.5%; Score 33; DB 16; Length 244;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVIGGK 7  
 1111111  
 Db 214 GTLGGK 220

RESULT 9  
 ID 09Y154 PRELIMINARY; PRT; 537 AA.  
 AC 09Y154;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE BCDNA:GH03163 PROTEIN.  
 GN BCDNA:GH03163 OR CG1408.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Anil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson S.R., Nixson K.A., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
 RA Aghayani A., Arcaluna T.T., Baxter E., Blazej R.G., Butenoff C.,  
 RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Flise E.,  
 RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
 RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Paclab J.M.,  
 RA Park S., Sequeira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,

RA Celniker S.E.;  
 RT "Full-length Drosophila melanogaster cDNA sequence.";  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A603775; AAF57089.1;  
 DR EMBL: AF145614; AAD38589.1;  
 DR Flybase: FBgn0027598; BCDNA:GH03163.  
 SO SEQUENCE 537 AA; 56128 MW; 704755F0092246B6 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 537;  
 Best Local Similarity 75.0%; Pred. No. 3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGGK 8  
 1111111  
 Db 175 GNLLGGK 182

RESULT 10  
 ID 09VA35 PRELIMINARY; PRT; 545 AA.  
 AC 09VA35;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE BCDNA:GH03163 OR CG1408.  
 GN BCDNA:GH03163 PROTEIN.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Anil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson S.R., Nixson K.A., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";



RL Science 287:2185-2195(2000).  
DR EMBL: AF003775; AAF57088.1; -  
DR FlyBase; FBgn0027598; BCDNA:GH03163  
SQ SEQUENCE 545 AA; 57302 MW; DBB92C9B979BF768 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 545;  
Best Local Similarity 75.0%; Pred. NO. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8  
|:|||||  
Db 175 GNLGSKK 182

## RESULT 11

ID 09PS81 PRELIMINARY; PRT; 607 AA.  
AC 09PS81:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MAD+PROTEIN(ADP-RIBOSYL)-TRANSFERASE, ADPRT  
OS Xenopus.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
OC Xenopodinae.  
OX NCBI\_TaxID=8353;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96007847; PubMed=7578427;  
RA Auer B., Flick K., Wang Z.Q., Haidacher D., Jager S., Berghammer H.,  
Kofler B., Schweiger M., Wagner E.F.;  
RA "On the biological role of the nuclear polymerizing NAD+; protein(ADP-  
RT ribosyl) transferase (ADPRT): ADPRT from Dictyostelium discoideum and  
RT inactivation of the ADPRT gene in the mouse.";  
RL Biochimie 77:444-449(1995).  
DR HSSP; P26446; 1A26.  
DR InterPro; IPR001357; BRCT.  
DR InterPro; IPR001290; PARP.  
DR InterPro; IPR004102; PARP\_reg.  
DR Pfam; PF00533; BRCT; 1.  
DR Pfam; PF00644; PARP; 1.  
DR Pfam; PF02877; PARP\_reg; 1.  
DR PROSITE; PS0172; BRCT; 1.  
SQ SEQUENCE 607 AA; 67496 MW; 54CDEBBE22079886 CRC64;

Query Match 80.5%; Score 33; DB 13; Length 607;  
Best Local Similarity 75.0%; Pred. NO. 3.5e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8  
|:|||||  
Db 185 GTVLGSKK 192

RESULT 12

ID 09C913 PRELIMINARY; PRT; 689 AA.  
AC 09C913:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHEICAL 77.5 KDA PROTEIN.  
GN F26A9.16.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
[1]  
RN RN  
RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome 1 BAC F26A9 genomic sequence";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC016163; AAG51819.1; -  
DR InterPro; IPR002885; PPR.  
DR Pfam; PF01535; PPR; 6.  
KW Hypothetical protein.  
SQ SEQUENCE 689 AA; 77486 MW; 16413082FF846D0D CRC64;

Query Match 80.5%; Score 33; DB 10; Length 689;  
Best Local Similarity 75.0%; Pred. NO. 4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8  
|:|||||  
Db 186 GTVLGSKK 193

## RESULT 13

ID 09EP75 PRELIMINARY; PRT; 64 AA.  
AC 09EP75:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE P0458A05.14 PROTEIN (B1157F09.4 PROTEIN).  
GN P0458A05.14 OR B1157F09.4.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Euphorbiaceae; Euphorbiaceae; Oryza.  
OX NCBI\_TaxID=4530;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
RT clone:P0458A05.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
RT clone:B1157F09.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP002870; BAB19405.1; -  
DR EMBL: AP003207; BAB64079.1; -  
SQ SEQUENCE 64 AA; 7300 MW; 269844DAF129933B CRC64;

Query Match 78.0%; Score 32; DB 10; Length 64;  
Best Local Similarity 75.0%; Pred. NO. 48;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8  
|:|||||  
Db 53 GTVLGSKK 60

RESULT 14

ID 0923G6 PRELIMINARY; PRT; 153 AA.  
AC 0923G6:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PROGESTERONE RECEPTOR (FRAGMENT).  
GN PR.

Search completed: July 30, 2002, 15:37:28  
 Job time: 333 sec

OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 ON NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang X., Das S.K., Paria B.C.;  
 RT "Mesocricetus auratus progesterone receptor (PR) mRNA."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY040534; AAK7950.1; -.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 153 153  
 SQ SEQUENCE 153 AA; 16909 MM; CQF00BF4ED3B8D95 CRC64;

Query Match 78.0%; Score 32; DB 11; Length 153;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGKK 8  
 | | | | | | | |  
 DB 87 GTVLGKK 94

## RESULT 15

Q9WHH2 PRELIMINARY; PRT; 174 AA.  
 AC Q9WHH2;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE M MATRIX PROTEIN.  
 OS Porcine reproductive and respiratory syndrome virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Niovirales;  
 OC Arteriviridae; Arterivirus.  
 OX NCBI\_TaxID=28344;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CH-1A;  
 RA Tong G.-Z., Qiu H.-J., Zhou Y.-J., Guo B.-Q., Zhang S.-J., Wang L.,  
 CAI X.-H.;  
 RT "Molecular Analysis of Genes encoding for the Structural Proteins of  
 RT Porcine Reproductive and Respiratory Syndrome Virus Strain CH-1a."  
 RL Ziran kexue jinzhan 10:147-153(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CH-1A;  
 RA Tong G.-Z., Qiu H.-J., Liu G.-Q., Xue Q., Zhou Y.-J., Qian P.,  
 Guo B.-Q., Liu B.-Q.;  
 RT "Complete Genomic Sequence of Porcine Reproductive and Respiratory  
 RT Syndrome Virus Strain CH-1a."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY032626; AAK44217.1; -.  
 DR InterPro; IPR001332; Arter1\_glycop.  
 DR InterPro; IPR003240; Porcine\_RR\_virus\_M.  
 DR Pfam; PF00951; Arter1\_glycop; 1.  
 DR Prodom; PD002878; Porcine\_RR\_virus\_M; 1.  
 KW Matrix protein.  
 SQ SEQUENCE 174 AA; 19011 MM; 69B81EBB511712BB CRC64;

Query Match 78.0%; Score 32; DB 12; Length 174;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGKK 8  
 | | | | | | | |  
 DB 153 GTVLGKK 160

Wed Jul 31 07:54:09 2002

us-09-813-214a-2.rsp

2

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:12 ; Search time 82.85 Seconds  
(without alignments)  
2844.883 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708  
Sequence: 1 MNHIVYVIFNKATGTFMAYV.....NCSADTQGHVCAVAGAFHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10697.5	99.9	2123	22	AAE00701
2	7380.5	68.9	2047	22	AAB69134
3	7357.5	68.7	2053	22	AAB69135
4	7113.5	66.4	1992	22	AAB69133
5	7110.5	66.4	1992	17	AAW04505
6	7110.5	66.4	1992	22	AAB69137
7	6600	61.6	2314	22	AAB69136
8	1185.5	11.1	2353	17	AAR93933
9	1169.5	10.9	2411	21	AAB23860
10	1063	9.9	1778	22	ABBS2677
11	1018.5	9.5	1104	21	AAB23856

12	1018.5	9.5	1104	21	AAB23859	Haemophilus influe
13	974	9.1	1004	21	AAB23857	Haemophilus influe
14	944.5	8.8	1002	21	AAB23854	Haemophilus influe
15	809.5	7.6	2039	19	AAW56322	Haemophilus paraga
16	723	6.8	2042	19	AAW56319	Haemophilus paraga
17	693	6.5	1094	21	AAB23858	Haemophilus influe
18	681	6.4	1098	17	AAR93932	Haemophilus adhesi
19	609.5	5.7	2344	22	AAU37120	Staphylococcus aur
20	601	5.6	5795	22	AAU37017	Staphylococcus aur
21	585.5	5.6	6281	22	AAU37403	Staphylococcus aur
22	582.5	5.4	679	17	AAR93934	Haemophilus adhesi
23	582.5	5.4	679	21	AAB23855	Haemophilus influe
24	550	5.1	2086	22	AAU34143	Staphylococcus aur
25	549.5	5.1	2586	22	ABBS6878	Drosophila melanog
26	543	5.1	2434	22	AAU34339	Staphylococcus aur
27	539.5	5.0	2599	21	AAU75098	Neisseria meningit
28	537	5.0	1601	18	AAW30292	Non-tyeable Haemo
29	533.5	5.0	2732	22	ABBS2855	Escherichia coli p
30	528.5	4.9	1598	18	AAW30291	Non-tyeable Haemo
31	522.5	4.9	2478	22	AAU34320	Staphylococcus aur
32	522.5	4.9	2478	22	AAU37374	Staphylococcus aur
33	515	4.8	1536	14	AAU41725	High molecular wei
34	513	4.8	2514	21	AAU75097	Neisseria meningit
35	512	4.8	1536	15	AAB63505	Haemophilus high m
36	512	4.8	1536	21	AAB01846	Haemophilus influe
37	510.5	4.8	2368	22	AAU34139	Staphylococcus aur
38	510.5	4.8	2368	22	AAU36796	Staphylococcus aur
39	509	4.8	1536	14	AAU41723	High molecular wei
40	508	4.7	1536	18	AAW30293	Non-tyeable Haemo
41	508	4.7	2383	21	AAB15945	E. coli proliferat
42	497	4.6	1529	14	AAU41732	High molecular wei
43	496.5	4.6	2048	21	AAU75096	Neisseria gonorrhoe
44	490.5	4.6	1477	14	AAU41724	High molecular wei
45	478.5	4.5	1477	15	AAB63506	Haemophilus high m

#### ALIGNMENTS

RESULT 1	
ID AAE00701	standard; Protein: 2123 AA.
AC AAE00701	
DT 02-JUL-2001	(first entry)
DE Moraxella catarrhalis outer membrane protein-106 (OMP106).	
KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;	
KW bacterial infection; immunogen; cytotoxic; antibiotic;	
KW bacterium immunisation;	
OS Moraxella catarrhalis.	
XX US6214981-B1.	
PN 10-APR-2001.	
PD 12-NOV-1997;	97US-0968685.
PF 03-MAY-1996;	96US-0642712.
PR (ANTE-) ANTEX BIOLOGICS INC.	
PA Tucker K, Plosilla L, Tillman UF;	
PI WPI: 2001-281002/29;	
DR N-PSDB: AAD04029.	
XX Novel nucleotide sequences encoding Moraxella catarrhalis outer	
PT membrane protein-106 polypeptide; useful for diagnosis of bacterial	
PT infections and as vaccine against Moraxella catarrhalis infection of	

PT mammals  
XX  
PS Claim 7; Column 53-64; 49pp; English.  
XX  
CC The present sequence is haemagglutinating *Moraxella catarrhalis* outer  
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic  
CC and prophylactic vaccine against *M. catarrhalis* infections of mammals.  
CC It is used for diagnosis of bacterial infections and as reagents for  
CC clinical or medical diagnosis of *M. catarrhalis* infections and for  
CC scientific research on the properties of pathogenicity, virulence and  
CC infectivity of *M. catarrhalis*. It is also used as a probe to identify  
CC the presence of *M. catarrhalis* in biological specimens and to identify  
CC other bacteria that encode a polypeptide related to *M. catarrhalis*  
CC OMP106. OMP106-derived polypeptides are used as ligands to detect  
CC antibodies elicited in response to *M. catarrhalis* infections and also  
CC as immunogens for inducing *M. catarrhalis*-specific antibodies which are  
CC useful in immunoassays to detect *M. catarrhalis* in biological specimens  
CC Cytotoxic antibodies are useful in passive immunisations against  
CC *M. catarrhalis*.  
XX  
XX Sequence 2123 AA;  
SQ

Query Match	99.98;	Score 10697.5;	DB 22;	Length 2123;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2122; Conservative	0;	Mismatches	0;	Indels 1; Gaps 1

Oy	1	MNHITKVENKATGPEMVAIVAYAKASHSHSGGSCANTQGVSRITLSPRIALAVLIGAT	60
Db	1	mnhitkvenkatgpemvaiwayakashshsggscatqgvsriltspariaalavligat	60
Oy	61	LNGSAVAGIGISEADCGKCGKANARGKDSIALGIDIAOLAGSOSIALGINKITVHNSNNNANI	120
Db	61	lngsavagigiseadcgkcgkanargkdsialgidialagsgsialginkitvhnsmnnani	120
Oy	121	CAKASGNSTIALGVDVLASGASHAISGDDLYLKKEYOQISELPIITROKALNDIYOL	180
Db	121	gakasgnestialgvdvlasgashaisgsddlylkketvqisellpiirgqkalndiyql	180
Oy	181	ADTNLOKRRFTHAOGHASTAVGANSYAKGHSNKGPMFATTEGNYSLAVGLTATKAKASS	240
Db	181	adtnlqkrrfthaghnastavgansyakghsnkgfmatteagtyslavglatakaass	240
Oy	241	IAVGSMAGAIGFPAATVAGSGSTQVNLNRGIALGFGSOVLQKNDVNVANVRAVAPDDQPI	300
Db	241	iavsgmagaiGFpaatvAGsgstQvnlNRgialGfgsoVLqkndvNVanvRAvAPddqpi	300
Oy	301	DNRKATPTKNCATDVFSIGNSNGNDSTRRKILINYGASADTDAYNAVQOLKAVYLANROI	360
Db	301	dnrkatptkncatdvfsignsgndstrrkilinygasadtdaynavqlkavylanroi	360
Oy	361	TFKGDDSNRRYKKGIGKRLITTTGGAQOISALTDHNIQVONDDGLYOLAERTLSLKRYVT	420
Db	361	tfkgddsnrryKkgigKrlITttGgaQoisALtdhniQvondDglYolaertLSlkryvt	420
Oy	421	ENLTANEKVTVGKTRLTTDKIGFPNDMNGIDESKRYPYDKDGIHAGOKITRKLTAAGVDD	480
Db	421	enltanekvTVgKtrLTtdKigFPndMngIdesKryPYdkdGIhagokITrklTAgvdd	480
Oy	481	DAATYGOLKATVQNAEESLQTFYTKYKDDKNGDANDSKITTVGKNKRPDGIQVNTLKLK	540
Db	481	daatygoLkAtVqnaEesLqTFyTKyKddKngDandSkITTVgKnKrpDgiQvntLklk	540
Oy	541	ENGVDVETTEGTVTFGLNONGNLGVSTLNNNGLSVKNTNSKOJQVGDGTTFPDIS	600
Db	541	engvdvetteGtvTFglNongNlGvstLnnngLsvkntnsKojQvgdgTTFpdis	600
Oy	601	NSKPGAGIENTTRITRDGIGFANNTGSLDANKPRLPTGINAGKELTNVOSAINPATNG	660
Db	601	nskpgagienttrITrdgIGfanntGslDankPrlPtgInagKeltnvosaInPatng	660
Oy	661	GOLDFMNLSTANNEKSSASATINDYVNLSDVPLTPAGDGPNTYTKLGLLKYKGGKT	720
Db	661	goldfMnlstanneKssasATindYvnlSDvPltpAGdGPntYtkLglLkyKggkt	720

Dd	661	gqltdmrlsiantekcksaaatlkdylnlsvqrp1tfaegtqgnvckkigbelkvygkct	720
Qy	721	ADDLTKNNIGVAVADSTONSLTVKIAKTLSDLAVDNTRKTTLASDQKTVDSGNNTAKLQND	780
Dd	721	addltknnlgyvadsldns1tvlk1tclsdldavnc1tlaaskvrlvdsqnn1tak1qngd	780
Qy	781	LTFSKONNGAIPARNKSTI-CVGS1KRTDNNG1ALDGGTYITPKDKYGFAPKOGS1DKSKP	839
Dd	781	ltfsqnggaiparnskt1iyvgd1k1tldnng1alqgt1y1tckdvgyfakqgs1dkskp	840
Qy	840	YLDKDKLKVESEVETTTMGINAGGKATIGLSNLTLDATNATTCVHTD1G1VDS2DKTPRAAS	899
Dd	841	yldkdklkvsevetl1tmginaaggkatiglsn1tldatnat1tghv1q1vg1ds2dktpraas	900
Qy	900	IGDVLNAGFN1KNNGDADFPVSTYD1VDF1NGNATTAKYTDGKASVAYDVYVNDGTTIH	959
Dd	901	igdv1nagfn1knnagdakdfvsvtyd1vdf1ngnat1takvtdgkaekvaydvvndgt1ih	960
Qy	960	LTGADGKKNQ1GVVTTTLTUTDAGDKGKA1NFVSNSGDKAL1NAK1DAN1MTLAGELRN	1019
Dd	961	ltgagdnqng1gvvt1tltctkdagdkka1nfvnsgdckal1nak1dan1mtlagelrn	1020
Qy	1020	TKGADVP1LQFQYKVKYKENGDDNDNDADTTVCBKDKNTQVMTLK1KGKNG1D1OTNMKG	1079
Dd	1021	tkgadad1qlf1qykvkvyengddndad1t1vckdkntqvn1lk1kgkng1d1otnmkg	1080
Qy	1080	TVTFG1INTOS1LAKGNMTT1NNNGLS1KNTAGNEQ1QVQADGVKFAKVNNGVAG1DGT	1139
Dd	1081	tvtf1gntqsg1kqgmnt1l1nnng1slkntagneq1qvgadgvkfakvngvvgag1dgt	1140
Qy	1140	TRTRIDELGFPGTNGS1DKSKPHLSKDGI1NAGKRTT1N1OSGEL1AONS1DAVTCGR1YDL	1199
Dd	1141	trtr1del1gfpgtngs1dkskphlskdg1naggkrtt1n1osg1el1aons1davtcgr1ydl	1200
Qy	1200	KTELENK1SS1RAK1AONS1LEEF5VADQGN1FVSP5SYDPSK1SDV1TFAENG1GTT	1259
Dd	1201	ktelenk1ss1tak1aons1leef5vadqgn1fvsps5sydpsk1sdv1t1aeng1glt	1260
Qy	1260	KVNGKVVAVG1DQ1FKG1TTRK1LTVGNNGKGI1YDSONGONT1TGLSNT1AVNTD1KCSV	1319
Dd	1261	kvnkgvvavrg1dq1fkg1ttrk1l1tvgnngkg1y1dsongont1tgl1snt1avntd1kcsv	1320
Qy	1320	RTTEOGK1IKDEDKTRAS1VDVLSAGFN1QNGEAVDFVSTYD1VNFADGNTTAKYTY	1379
Dd	1321	r1teegk1ikdedk1traas1vdv1sagfn1qngaeavdfvsvtyd1vnf1adgn1t1ak1yty	1380
Qy	1380	DDTSKTSVVYVDVWDDTTL1EVDKDK1KGYVTTTLT1STG1GANKFAL1SNOATG2DAL1VKRS	1439
Dd	1381	ddtsktsvvvydvvd1t1levdkdk1gyvt1t1t1stg1qan1fal1snga1tgdal1vksd	1440
Qy	1440	IVAH1NLT1SGD1OTKAKASQANSAG1VADGKRV1YDSTDNKRY1YQAKNDG1VDP1TKEVA	1499
Dd	1441	1vah1nlt1sgd1otkakaasqansag1vadaqkv1ydstdnk1yqakndg1vdp1tkeva	1500
Qy	1500	KDK1LVAQAQ1PDG1L1AOMNYSV1NKEQVNDANKKOG1INEDNA1FVKGLEKAA1SDNKTRNA	1559
Dd	1501	kdk1lvaqaq1pdg1l1aomnyksv1nkeqvndankkqg1inedna1fvkglekaasdnk1tka	1560
Qy	1560	AVT1GDLN1AVAQ1PFL1FASGNTTAKK1GELT1IKSGOD1D1TK1D1NN1GVYAG1D1GFTV	1619
Dd	1561	av1tvgd1n1avaq1p1fl1taqdt1gt1ak1g1tcl1k1qg1d1ct1dn11g1vva1q1d1gt1v	1620
Qy	1620	K1A1DL1TN1NSVNA1GTR1IDDK1GVSEF1DSS1GQAKKANT1P1LSANG1DL1GK1V1SNV1SGKTK	1679
Dd	1621	k1a1d1cn1nsvnaag1k1iddk1gsvf1dss1gqakant1p1lsang1dl1gk1v1snvg1k1t	1680
Qy	1680	DTDAAN1VOQLNEVRN1L1GLGNAGDNDA1GNOVN1IAD1KKDPNSGSS1NFR1V1KACT1VLGG	1739
Dd	1681	dt1daan1vg1l1evrn11g1gnagndn1ad1g1nva1ad1kk1dpns1gss1nfr1v1kag1vt1gg	1740
Qy	1740	KGNNDTEK1ATG1QVQV1VDK1DGNAND1LS1VWVW1K1TOKD1SK1ALLAT1YAA1QOT1V1LNN	1799
Dd	1741	kgnndte1at1g1v1qv1vd1k1dgnand1ls1v1wv1k1t1okd1sk1all1at1yaa1q1ot1v1l1nn	1800

CC 1800 PAEATDRINEGGIRFFHNDNOEPYVGRGICDSSASGKHSVAIGROAKDGEAAVAIG 1859  
 CC |||||||  
 CC 1801 paeatdrineggirffhndnqepvgrngidsaaqkhsvalgfyakadgaavaig 1860  
 CC |||||||  
 CC 1860 ROTOGNOSIAIGDMAATGQOSIAIGTGNVAGKHSAGIDPSTVKADNCSYVGNNOF 1919  
 CC |||||||  
 CC 1861 rltqagngsialgdnaqatqdsiaigtgnvvaqkhsaigdpstvkaadnsysvgnmqf 1920  
 CC |||||||  
 CC 1920 TDAQTDFEGVGNNTVTESNSVALGSNSAISAGTHAGTOAKKSDGTAGTTTATGATGV 1979  
 CC |||||||  
 CC 1921 tdaqtdfvgygnntvtresnsvalgsnsaisagthagtdqkksadgtagtttagatgtv 1980  
 CC |||||||  
 CC 1980 KGFAGQTVAGVAGSASAEKRIQVNAAGEVATSTDAVNSQLYKATQIGANNINELDH 2039  
 CC |||||||  
 CC 1981 kfgagqtvagvagsaqaerriqvaagevatsldavpysqlykatqiganaetneldh 2040  
 CC |||||||  
 CC 2040 RIHONENKANGISSAMAMPAQVIGRSMVTGATIHNGOGAVAVGLSLDNGOMV 2099  
 CC |||||||  
 CC 2041 rihonenkanagissamamampqaylpgtismvtglatihnggavavglslsdngqvw 2100  
 CC |||||||  
 CC QY 2100 FKINGSADTQGHVGAAGAGFHF 2122  
 CC |||||||  
 CC Db 2101 fkingadtdqghvgaagagfhhf 2123  
 CC |||||||  
 CC RESULT 2  
 CC AAB69134  
 CC ID AAB69134 standard; Protein: 2047 AA.  
 CC |||||||  
 CC AC AAB69134;  
 CC |||||||  
 CC DT 24-APR-2001 (first entry)  
 CC |||||||  
 CC DE M. catarrhalis strain 4223 genomic 200kda protein seq ID NO:7.  
 CC |||||||  
 CC KM Moraxella catarrhalis strain 4223; major outer membrane protein;  
 CC 200kda outer membrane protein; antibacterial; immunogenic; infection;  
 CC otitis media; detection.  
 CC |||||||  
 CC OS Moraxella catarrhalis.  
 CC |||||||  
 CC PN MO200107619-A1.  
 CC |||||||  
 CC PD 01-FEB-2001.  
 CC |||||||  
 CC PF 26-JUL-2000; 2000WO-CA00870.  
 CC |||||||  
 CC PR 27-JUL-1999; 99US-0361619.  
 CC |||||||  
 CC PA (CONN-) CONNAUGHT LAB LTD.  
 CC |||||||  
 CC PI Loosmore SM, Sasaki K, Yang Y, Klein MH;  
 CC |||||||  
 CC DR WPI: 2001-159722/16.  
 CC |||||||  
 CC DR N-PEDB; AAF59102; AAF59103.  
 CC |||||||  
 CC PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,  
 CC useful in protective vaccines and for diagnosis  
 CC |||||||  
 CC PS Claim 1: Fig 3A-W; 247pp; English.  
 CC |||||||  
 CC XX The present invention describes an isolated and purified nucleic acid (I)  
 CC that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kda outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines. (II), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II) and its  
 CC fragments are used as probes for identifying/cloning 200 kda protein  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis strain 4223 genomic 200kda protein, which is given in the  
 CC exemplification of the present invention.  
 CC XX  
 CC Sequence 2047 AA:  
 CC  
 CC Query Match 68.9%; Score 7380.5; DB 22; Length 2047;  
 CC Best Local Similarity 68.3%; Pred. No. 0;  
 CC Matches 1557; Conservative 117; Mismatches 217; Indels 387; Gaps 21;  
 CC  
 CC 1 MNHLYKVFENKATGTFMAVAEYAKSHSTGGSCATGVSFRTPSFARIALAVLIGAT 60  
 CC |||||||  
 CC Db 1 mnhlykvlfnkaqtlmaevaeyakshstggscatgvsfcslfarialavlyigat 60  
 CC |||||||  
 CC QY 61 LMSGSAV-----GIGISEADGCKGANARDKSIAIGDIAQALGOSIAIGDKIYH 112  
 CC |||||||  
 CC Db 61 lsgsayaqktdkthialgeqngprsgtakadgdralaigenanagqgaiaigsnkty 120  
 CC |||||||  
 CC QY 113 NSNNNANIGAKASGNEISIAIGDVLASGHSIAIGSDLYL-----KKEYVOQISELL 165  
 CC |||||||  
 CC Db 121 ngsldkigtatqgesiaigdvkasgdsiaigsddhlldqhnpkhpkytllndi- 179  
 CC |||||||  
 CC QY 166 PIIRGOKALNDIYQIADPTNLQKYRRTTHAGHASTAVGAMSYAKGHSNAGTATAGTY 225  
 CC |||||||  
 CC Db 180 --inghavlkelfrskndv-kyrrtlasgnastavaamsyagqghfnagfratatsay 236  
 CC |||||||  
 CC QY 226 SLAVGLTATKAASSIAVGSNAQIIGFAATAVAGSTQVNLNRGIALFGSQVLQKNDVN 285  
 CC |||||||  
 CC Db 237 slavglataegqslaisgdatssslgalaiaqgrtaqgslaiqgsavtgsbn--- 293  
 CC |||||||  
 CC QY 286 AANVR-AVAPDDNOPIDNRKATFKAGATDFESIGNSNGDSIRKRIYNGASADTDAY 344  
 CC |||||||  
 CC Db 294 --nsrpaylp-nqcaldpkfgatnntkagpl-----sigsnlkrlinvagavntkdav 345  
 CC |||||||  
 CC QY 345 NVAQLEKAVKLA-NROTFFGDDSNNRVEKGLKTLITTGAGTSLDTHNIGVONGD- 402  
 CC |||||||  
 CC Db 346 nvaqlaavvkwakerrlftfgddnstavkigldntlikgsaetnaltdmngvveadn 405  
 CC |||||||  
 CC QY 403 -GLKVOLEETLSLKAVTNTENLTPANEKVTGK----- 433  
 CC |||||||  
 CC Db 406 sgllkvklaktlnlntevnttlaattvkysssstlaelsalfttgnhtsgstskt 465  
 CC |||||||  
 CC QY 434 -----TRLTDDKIGFTNDMNCIDSKPYLDK-----D 460  
 CC |||||||  
 CC Db 456 vygvngvktfnnaettaaigttrlttrdkigfardgdvdeqapyldkqjkygsvaltid 525  
 CC |||||||  
 CC QY 461 TGIHAGGOKITKLAVGVDDDAATYQOLK----- 489  
 CC |||||||  
 CC Db 526 ngldagnkkslnlaksandavltleqkaakprlnagaglsyvtpteisvdaksnvtap 585  
 CC |||||||  
 CC QY 490 -----RVNQTAESALOTFTYKVY 507  
 CC |||||||  
 CC Db 586 tynlgvkttelnsdgtadsktsvksqgtnslvtaehlasylneutrtdaalsqfve- 644  
 CC |||||||  
 CC QY 503 DKNGNDANDSKITIVGNKPKPDGTQVNTFLKKGNGGVDTTETNGTVFELNONGNLTGV 567  
 CC |||||||  
 CC Db 645 -----edddanaalvaktctknagavsilklkygngltvaktkdytvlfiglsqdsqllig 700  
 CC |||||||  
 CC QY 568 NSTLNDGLSVKNTNSMKOYGVADGTFITPDISKRSKAGIENRTTRRGIGFANNIGS 627  
 CC |||||||  
 CC Db 701 kstlndglvtvkd-pegdvgaangikflnvngsnpgtqjanarlttr----- 747  
 CC |||||||  
 CC QY 628 LDANKPRLTPTGINAGKEKELTNVOSAINPATNGQLDFMRLSTANTEKSGSATIKDLY 687  
 CC |||||||  
 CC Db 748 ----- 747  
 CC |||||||  
 CC QY 688 NLSQVPLTFAGDTGPNVTKLIGELIKVKGKTTADLTKNNGIGVADSTDSNLTWLAKT 747  
 CC |||||||  
 CC Db 748 ----- 747  
 CC |||||||  
 CC QY 748 LSDDAVNTKTLTASDKVTVDSGNNNTAKLQNGDLTFESKQNTGATPATNSKTIGVDGLKFT 807

```

Db 748 ----- 747
QY 808 DNNGIALDGTYYITKDVYFPAKQDSKSPYIDKDKLKVGEVETTINGINAGKAITG 867
Db 748 -----dkigtsgdgaavdltnkpyldgdklgygnakltnglinsgkaltg 792
QY 868 LSNLTLDATNATGHTVQOLGVISTDSTDKTAAASIGDVLNAGNFKNNGDADDEVSTYVTD 927
Db 793 Lspilpsiadqsrnielngitgdkksnaasindlntgfnlknmnpidfvsydlavd 852
QY 928 FINGNATTAAYYD - GKASRAVDVNDGTTHLVTGADGNKNOIGVKTTLTKTDKAGD 985
Db 853 fangnatlatvthdntsksvydvndvdtlltltgtdnk - klygkttklktsangn 911
QY 986 KATFVSNSGDALKINAKIADNLNLTAGEIRTKGTATATOTFOYKVKKEKGGDDND 1045
Db 912 tatfnvnsdcdalvlnakdiaenlntlakeihltkgtadaltqftvckvdenmad-d 970
QY 1046 ADTITVCKDAKTNOVNTLTKLKGKNGLDIOTNKDGTVEGINTOSGLKAGNNTTLNNGLS 1105
Db 971 analvtvgknaangvntltlkgenglnlktokngvtlfglntsglkaq - kstlndgals 1029
QY 1106 IKNTAGNEQIQVADGVKFAKV - NNGVVGAGIDGTRIRTRDEIGFAGTNGSLDKSPHLS 1164
Db 1030 iknptgseqlvgadgvykfakvnmngvvgagldgttrlttrdelgltgtnsgldkskphls 1089
QY 1165 KDGINAGKRTTNGSEIATONSMDATVGTGKIYDKTELENKTSSTAKTQNSLHESVA 1224
Db 1090 kdglnaagkkltnlqsgelqnsndavtggklydlkleenklstaktaqnslhetsva 1149
QY 1225 DEQGNFTVSNPYSSTYDTSKTSVDYTFPAGENGITTKVKNKGVNRVIGDOTGTLTPKLTVG 1284
Db 1150 deggnltfvenpysytsktsdvtlftagengitkvnkvvrvvgldgtqtltpkltvg 1209
QY 1285 NNNKGIVIDSQNCQNTITGLSNTLANVTNDKGSVRTTEGGKIIKDEKTRAASIVDLS 1344
Db 1210 nnnkgividsqngntltglntlanvtnkgsvrtteggntlkdedktraasivdls 1269
QY 1345 AGFNLQNGEAVDVFSTYDVPVNFADGNATKATYTDSTKTSVYVYVNDVDTTIEKDK 1404
Db 1270 agfnlqngaeavdvtvsydvtnladgnatcavkcyddtsktsvvydvndvdtliekdk 1329
QY 1405 KLGKVTTLTSTGTGANKFALSNOATGDLVKAADIVAHNLTSGDIOTAKGASQANSSA 1464
Db 1330 klgkvtltltstgtgankfalsngatgdalvksadivahntlsgdlqtakgsganssa 1389
QY 1465 GYVDADGNKVIYDSTDNKYYIOAKNDGTVDKTEKVAKDVLAAQOTPDGTLAQMNVKSVIN 1524
Db 1390 gyvadagnkviydstdnkyyqakndgtvdktkvakdklvaqagtpdgtlaqnmvksvin 1449
QY 1525 KEQVNDANKKOGINEDNAFPGLEKASDNKTKNAATVGDNLAAVQOTPLTFAGDGTGTA 1584
Db 1450 keqvndankkgyinednafpgylekaasdnktknaavtvgdnlavagqtpcltfagdggtta 1509
QY 1585 KKLGETLTIKGGOTDTNKLTDNNINIGVAGTGTGVLAKDLTLNLSNAGGTIKIDKGSV 1644
Db 1510 kklgetltikggotdtnkltdnnlignvagtgtvtvklakdltnlnsnaggtikidkgsv 1569
QY 1645 FVDSGGAKANTPYLASANGDLGKAYISNKGKTKOTDANAVOOLEVNRNLGLNAGND 1704
Db 1570 fvdssggaakanpylasangldlgkaylsnvkgtktdaanavglnevrnlglagnagd 1629
QY 1705 NADONOVNIDIKKDPSSGSSSNRTYKAGTVGCGKGNDEKLANGVGVGVGDKXGNN 1764
Db 1630 nadonovnidikkdpssgsssnrtvklagvlggkgnndeklanvgvlgvdkkgnn 1689
QY 1765 GDSNVWVKTQKDSKKALATYNAAGOTVYLTNNPNAEALIDRIINEGIRFEHNDGNOEP 1824
Db 1690 gdsnvwvktqkdskkallatynaagqtnyltnnpaealidrinegirfehndngnpe 1749
QY 1825 VVGGRNGIDSSAGKHSVAIGFOAKADGEAANAIGROFOAGNOSIAIGDNAQTGDSIA 1884

```

```

Db 1750 yvgrngidsaasgkhsvalfgakadgaeavaigrqtgagngslagdnaqtgdsia 1809
QY 1885 IGTGNVYAKHSGAIGDEBETVAKDNSYSGNNNOFTDAQTQVFEVGNNTTYESNSVAL 1944
Db 1810 lgtgnvvaqkhsaglygdpstvkdnsysvgvnnqftdactqlvtlfgvgnltvtesnval 1869
QY 1945 GSNLSASGTHAGTOAKKSDGTAGTTTGTAGTGVKGAFGAGOTPAVAGAVSGASGARRION 2004
Db 1870 gnsalsagthgtagkksdgtagttttagatgvtkygfyagqtaavavsgasgearrion 1929
QY 2005 VAAGEVSATSDAVNSQLYKATQGIANATNELDRIHONENKANAGISSAMAMASMPQA 2064
Db 1930 vaagevsatcdavogsglykatgslanaetnelhribqnenkanagissamamasmpqa 1989
QY 2065 YTPGRSMVTGATIHNGQCAVAVAGLSKSDNCGWFKINGSADTJGHHVGAAGAFHR 2122
Db 1990 ytpgrsmvtglatlhngqgavavaglskdsngqwfklngsadtlghvgaagafhr 2047

RESULT 3
AAB69135
ID AAB69135 standard; Protein; 2053 AA.
XX
AC AAB69135;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 08 200kda protein SEQ ID NO:9.
XX
KW Moraxella catarrhalis strain 08; major outer membrane protein;
XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;
XX otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
PN WO200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
DR WPI: 2001-159722/16.
XX
N-PSDB: AAF59104.
XX
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX
PS Claim 1; Fig 4A-V; 247pp; English.
XX

The present invention describes an isolated and purified nucleic acid (I)
that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
The 200 kDa outer membrane protein (II) has antibacterial activity and
can be used in vaccines. (II), and its truncated versions, are used as
immunogenic compositions and vaccines to protect against M. catarrhalis
infections, particularly otitis media in humans. (II) is also used as
an antigen in immunoassays for detecting specific antibodies (Ab), and to
generate Ab. (I) are used for recombinant production of (II) and its
fragments are used as probes for identifying/cloning 200 kDa protein
genes from other strains, and for diagnostic detection of M. catarrhalis.
(I) makes possible production of large amount of recombinant immunogens.
Expression of truncated versions of (II) reduces toxicity of the protein
towards the Escherichia coli host. The present sequence represents the
M. catarrhalis strain 08 200kda protein, which is given in the
exemplification of the present invention.
XX
Sequence 2053 AA;

```



Query Match 68.7% Score 7357.5 DB 22 Length 2053;  
 Best Local Similarity 70.9%; Pred. No. 0;  
 Matches 1574; Conservative 117; Mismatches 264; Indels 265; Gaps 35;

QY 1 MNHIVYFNKATGTGMAVAEYAKSHSTGSCATGQVSGVRLTSPARIALAVLYGAT 60  
 1 mnhivylfknatgtgmaaveayakshstgsgcatgqvsgvrltsparialavlygat 60  
 DB 1 mnhivylfknatgtgmaaveayakshstgsgcatgqvsgvrltsparialavlygat 60  
 QY 61 LNSAYAGIGISEADG-----KGANAKGDKSIAIGDIAQALGSSOIAIGDKIV 111  
 61 lnsayagigigiseadg-----kganakgdksiaigdiaqalgssoiaigdkiv 111  
 DB 61 lnsayagigigiseadg-----kganakgdksiaigdiaqalgssoiaigdkiv 111  
 QY 112 HNSNNNNNIGKAKSGNSIAGIDVYLASGHAISAIGSDDIYLLKKEPQQOISELLPIIRGQ 171  
 112 hnsnnnnnigkaksngsiagidvylasghaisaigsdidiylkkepqqoisesllpiirgq 171  
 DB 121 pnsngnsvghakngesialagdvlaegdaslaigsdidiylp-nldlknethkllhgh 179  
 QY 172 KALNDIOLADTNLOKRYRTHAOGHASTAVGAMSAKGFHSMFGRATAGTYSLAVGL 231  
 172 kalndioladtnlokryrthagoahastavgamsakgfhsmafgratagtyslavgl 231  
 DB 180 elkkltstvgki-kyrrtaaghaslavamsyagghhsnaigtvataeayslavgl 238  
 QY 232 TATKKAASSTAVGSAQAIGFATAVGSGTOVNLNGIALGFSQVLOKDNVANAANVRA 291  
 232 tatkkasstavgsaqaigfatavgsgtovnlngialgfsqvlokdndvnanavra 291  
 DB 239 aagatkgsslavgsnakanafaataaigntvnlgrvalgfsqvlldrdnntdas---a 295  
 QY 292 YAPDNDQPIDNRKATPKNGATDVESIGNSGND-STRRKINNGAGSADTDANVNAOLK 350  
 292 yapdndqipidnrkatpkngatdvessignsngnd-strrkinngagsadtdanvnaolk 350  
 DB 296 yvp-lykladqykatrgygdstdlfsignsnmslrrklingagsrdtdavnaavl 354  
 QY 351 EAVLANQOITFEKDDSDN--NRVEKGLGKTLTIGGAQTSALTDHNLGVONGGKLEVOL 408  
 351 eavlanqoitfekddsdn--nrvekglgkltltiggaqtsaltdhnlgvonggklevol 408  
 DB 355 lveelanrkitfkgdgnmsnsverglngltlkgdgaqnaiteanlgvntdnglkvkl 414  
 QY 409 AETLSKMTTEN-----LT-----AN 426  
 409 aetlskmtten-----lt-----an 426  
 DB 415 akeltgltsvatnkitvsnntnnnaelqsggltfspitgktdktvysldglkftndsn 474  
 QY 427 EKVAVGKRLTLDKIGFTMDNGIDESKRPYDKDTGTHAGGQKTIKTLAGVDDADATYG 486  
 427 ekvavgkrltldkigftmdngideskrpydkdtgthaggqktiktlagvddadatyg 486  
 DB 475 slakgytrllkklkgtaglndgydeskpyld----- 506  
 QY 487 QLRKVNQTAESALQTFYKVKVDKNGNDANDSKITTVGKNKRPDQYVNTLKGENGVDV 546  
 487 qlrkvnqtaesalqtfykvkvdkngndandskittvgknkrpdyvntlkgengvdv 546  
 DB 507 ----- 506  
 QY 547 TTEINGYTFELNONGLTGNGSTLNDGSLVKNNTNSKQIQVAGADITFTDTSKRPG- 605  
 547 tteingytfelnongltgngstlndgslvknntnskqiqvagaditftdtskrpg- 605  
 DB 507 -----nekikvgnstlinsgltvnttgnkqiyganglkatvannant 552  
 QY 606 AGIENTRITFDGIGFANNGTSLDANKPRL-----TPGTNGGKRELTVVQS 652  
 606 agientritfdgigfanngtsgldankprl-----tpgtnggkreltvvqs 652  
 DB 553 satvglatieekigfagndgvdagpyldkerlkvgrveitdsjngnhtlgl-- 610  
 QY 653 AINPATNGQGLDFMNRSLSTANTEKSGSATIKDLYNLQVLFPAAGTGPVWTKLGEI- 711  
 653 ainpatngqgldfmnrslstanteeksgsatikdlynlsqvlfpagtgtpvwtklgei- 711  
 DB 611 -----tng-----iantd-----avtlkqklkdak-pltlaagd-gslinsngdlv 648  
 QY 712 -----LKVGKGTATDADLTKNNGVADSDNSITVRLAKTLSD-LDAVNTKT 758  
 712 -----lkvgkgtatdadtlnngvadsdntsvrlaktlslsd-ldavntkt 758  
 DB 649 dsagntitpnylsvtktlinsngtsgnktfsvanhnglsv-takdladylkvne-- 705  
 QY 759 LTAASKYVVDGSGNNTAKIQNDGLTFPSKONTGATPA--TNSKTITGVDGLKFTDNNGIALDG 816  
 759 ltaaskyvvdsgnntakiqndgltfpskontgatpa--tnsktitgvdglkftdnnngialdg 816  
 DB 706 -----tadaalpsfkvngd-----nsnaalvgktdngktfnt--lkkgengvnl-- 750  
 QY 817 TTYITKDKVGAKEGSDSKRPYLKDKLKVGEVETITNGI-----NAGGA 864  
 817 ttyitkdkvgaakegsgskrpylkdklkvgevetitngi-----nagga 864  
 DB 751 ttnratgvtl-----gidsgng-lltpklvgs-----dtgnrnlviegypsaadnsgtknl 801  
 QY 865 ITGSLNTLTDATNMTGHTVLOGLIVSDTKTAAISIGDVLNAGFNKNGAKDFVSTYD 924  
 865 itgslntltdatnmtghtvlgvldstktaaaisigdvlnagfnknkgakdfvstyd 924  
 DB 802 lkgslptlpslaspsgrnialgnltteekdkenaasiddvlnagfnlknngkdkdfvstyd 861

QY 925 TVDFINGNATKATKYVD--GKASKVAYDVNVGDGTTIHLTGADGNKNOIGVTKTTLTKDA 982  
 925 tvdfingnatkakyvd--gkaskvaydvnvgdgttihltgadgnknnoigvtkttltkda 982  
 DB 862 tvdfingnatkakyvd--gkaskvaydvnvgdgttihltgadgnknnoigvtkttltkda 920  
 QY 983 KGDRAINSVSGDDKALINARDIADNLTLAGELRNTKGTADTALQTFQVKKVENGDD 1042  
 983 kgdrainsvsgddkalinaradiadnltlagelrntkgtadtalqtfqvkkvengdd 1042  
 DB 921 ngn-attfstcd--ddhalvkasdiagnltlaeelhtkytantalqtlftvkvdn-dk 976  
 QY 1043 DNDADITVYGKAKTQVNTLKLKNGSLDIOGTNDGVTPEINOSGKAKGNNTLNN 1102  
 1043 dndaditvygkaktqvntllklkngslديوگتندگوتپینوسگاکگننتلنن 1102  
 DB 977 addtnalvkgdgtsgkvntllkkgngldlkdkgvltglntqsgllkagdsctllmn 1036  
 QY 1103 GLSKRTAGNEQIOVGADGVKFAKVNNGVAGIDGTRITRDEIGFAGTNSLPSKPH 1162  
 1103 glskrtagneqiovgadgvkfakvnnvgagidgtritrdeigfagtntslpskph 1162  
 DB 1037 gslkrtagneqiovgadgvkfakvnnvgagidgtritrdeigfagtntslpskph 1096  
 QY 1097 lskglnagkkitnngsgeialkshdavlsgklydlktelenklsstaktagnshfs 1156  
 1097 lskglnagkkitnngsgeialkshdavlsgklydlktelenklsstaktagnshfs 1156  
 QY 1223 VADGNNFTVSNPYSYDTSKTSDVITFAGENGITTKVKNGVVRGIDOTGTLTPKLT 1282  
 1223 vadgnnftvsnpysydtskttdvltfagengittkvknvvrgidotgtltptklt 1282  
 DB 1157 yadegnnftvsnpysydtskttdvltfagengittkvknvvrgidotgtltptklt 1216  
 QY 1283 VGNNGKGIYIDSONCONITGLSNTLANVTNDKGSVFTTEGKIHKDDKTRAASIVDV 1342  
 1283 vgnngkgiyidsonconitglslntlanvtnndkgsvftteggikikddktraasivdv 1342  
 DB 1217 vgnngkgiyidsonconitglslntlanvtnndkgsvftteggikikddktraasivdv 1276  
 QY 1343 LSAFNLONGEAVADVSTYDVNFADGNATKATKYVDITSKTSKVVYDVNVDDTTIEVK 1402  
 1343 lsafnlongeavadvstydvnfnadgnatkakyvdiitksktskvvydvnvddttievk 1402  
 DB 1277 lsaflngngeavadvstydvnfnadgnatkakyvdiitksktskvvydvnvddttievk 1336  
 QY 1403 DKRLGYKTTTLTSTGKANKFALSNOATGDALYKADYVAHLNTLSGDIOTAKKASQAMS 1462  
 1403 dkrlgyktttlstgkankfalsnoatgdalylkadyvahlnltsgdiotakkasqams 1462  
 DB 1337 dkrlgyktttlstgkankfalsnoatgdalylkadyvahlnltsgdiotakkasqams 1396  
 QY 1463 SAGYVADGNKVIYDSDTKNRYOAKNDGYDKREKAKKILVAQAOTPGTLAOMNVKSV 1522  
 1463 sagyvadgnkvlydsdtknryoakndgydkrekakkilvagaotpgtlaomnvksv 1522  
 DB 1367 sagyvadgnkvlydsdtknryoakndgydkrekakkilvagaotpgtlaomnvksv 1456  
 QY 1523 INKEQVNDANKKOGINEDNAFVKLEKASDNKTKNAAVTVGDNLNAVAOTPLTFAGDTGT 1582  
 1523 inkeqvndankkoginednafvklekasdnktnaavtvgdnlnavaoptlftagdtgt 1582  
 DB 1457 inkeqvndankkoginednafvklekasdnktnaavtvgdnlnavaoptlftagdtgt 1516  
 QY 1583 TAKKIGETLIRKGGOTDKNLTDMNNGVAVAGTGFVYLAKDLTNTNSNAGGTXTDDKG 1642  
 1583 takkigetlirkggotdknltdmnngvavagtgvfylvlakdltntnsnaggtxtddkg 1642  
 DB 1517 takkigetlirkggotdknltdmnngvavagtgvfylvlakdltntnsnaggtxtddkg 1576  
 QY 1643 VSPFVDSGQAKANTPYLSANGDLGKGVTSNVGKGTKDPDAANVOOLNRYMLLGNAG 1702  
 1643 vspfvdsqakantpylsangdlgkgvtsnvvgkgtkdpdaanvoollnrymlngnag 1702  
 DB 1577 vspfvdsqakantpylsangdlgkgvtsnvvgkgtkdpdaanvoollnrymlngnag 1633  
 QY 1703 NDNAQGNQVNIADIKKDPNSGSSSNRTVIAKAGTVLGGKGNNDTEKLATSGVOYGVDDKDN 1762  
 1703 ndnaqgnqvniadikkdpnsgsssnrtvialkagtvlggkgnndteklatsgvoygvddkdn 1762  
 DB 1634 ndnaqgnqvniadikkdpnsgsssnrtvialkagtvlggkgnndteklatsgvoygvddkdn 1693  
 QY 1763 ANGDLSNWWYKTKDQSGKALLATTYAAAGOTNLTNNPAPADIRIEOGIRFPHVNDGNO 1822  
 1763 angdlsnwwytkdqsgkallattyaaagotnltnnpapadirieogirfphvndgno 1822  
 DB 1694 angdlsnwwytkdqsgkallattyaaagotnltnnpapadirieogirfphvndgno 1753  
 QY 1823 EPVVOGNGTIDSSASGHSVAIGFOAKADGEAVALIGRTOAGNOSTAIGDNNQOATGDS 1882  
 1823 epvvoengtidssasghsvaigfoakadgeavaligrtoagnostaigdnnoatgds 1882  
 DB 1754 epvvoengtidssasghsvaigfoakadgeavaligrtoagnostaigdnnoatgds 1813  
 QY 1883 IAICTGNVAVGKHSAGIGDPSTYKADNSYVGNNOFTDATOTDVGVGNNITVTSNSV 1942  
 1883 iaictgnvavgkhsagigdpstykadnsyvgnnofdtatodvvgnnitvtsnsv 1942  
 DB 1814 iaictgnvavgkhsagigdpstykadnsyvgnnofdtatodvvgnnitvtsnsv 1873  
 QY 1943 ALGNSAISAGTHAGTOAKKSDGTAGTTTACATGTVKGFAGOTAVGAYSVGASGAKERI 2002  
 1943 algnsaisagthagtoakksgdtagtttacctgvtvkfgagotavgaysvgasgakeri 2002  
 DB 1874 algnsaisagthagtoakksgdtagtttacctgvtvkfgagotavgaysvgasgakeri 1933  
 QY 2003 QNVAAGEVSATSDAVANGSLYKATOGIANATWELDHRILHONENKANAGISSAMAMASMP 2062

Db 1934 qnvaagevsatctdavnsgslykatqslanaetndhrhqnknkagissamamamp 1993  
 QY 2063 QAYTPGSRMVGATTHNGGOGAVANGLSKLSDNGOWFKINGSADTQGVGAAGAGPHF 2122  
 Db 1994 qaylpgtmsvvgglatthnggavavglsklsdnqgwfikingsadltghvgaavagfthf 2053

## RESULT 4

AAB69133  
 ID AAB69133 standard; Protein: 1992 AA.

XX AAB69133;

XX 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 lambdaDEMBL3 clone 200kDa protein SEQ ID NO:3.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosemore SM, Sasaki K, Yang Y, Klein MH;

XX WPI: 2001-159722/16.

DR N-PSDB: AAF59100, AAF59101.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis -

XX Example 3; Fig 2A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)  
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.  
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and  
 CC can be used in vaccines. (III), and its truncated versions, are used as  
 CC immunogenic compositions and vaccines to protect against M. catarrhalis  
 CC infections, particularly otitis media in humans. (II) is also used as  
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to  
 CC generate Ab. (I) are used for recombinant production of (II) and its  
 CC fragments are used as probes for identifying/cloning 200 kDa protein  
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.  
 CC (I) makes possible production of large amount of recombinant immunogens.  
 CC Expression of truncated versions of (II) reduces toxicity of the protein  
 CC towards the Escherichia coli host. The present sequence represents the  
 CC M. catarrhalis strain 4223 lambdaDEMBL3 clone 200kDa protein, which is  
 CC used in the exemplification of the present invention.

XX Sequence 1992 AA;

Query Match 67.4%; Score 7113.5; DB 22; Length 1992;

Best Local Similarity 67.6%; Pred No. 6.8e-296; Mismatches 216; Indels 387; Gaps 21;

Matches 1503; Conservative 117; Mismatches 216; Indels 387; Gaps 21;

QY 56 VIGATLNGSAYA-----GIGISEADGKGAGANRGSIAIGDIAQALGSQSTAIAGD 107

Db 1 vlgatlsagayagkdkthlaigeqnprsgtkadgdtralaiganaagsggalaigs 60

QY 108 KTVHNSNNANIGAKASGNESIAIGDVLAGSHASIAIGSDLYL-----KRETVQ 160

Db 61 snktvngssldkigtatqgesiaigdvkaasdasiaigsdhlldghnphkphkgtl 120  
 QY 161 ISELLPIIRGOKALNDIYQADTNLQKRRTHAOGHASTAVGAMSTAKHFSNATRTAT 220  
 Db 121 indl---inghvalkeirskndv--kyrrtlasgastavagmsyagghfanaigtlrat 176  
 QY 221 AEGTYSILAVGLTPTAKAASSIAVGSNAQALGFATVAGSGSTOVNLRGIALGSGSQLVK 280  
 Db 177 aksaayslaavglaaeagqsliaigsdatslgaiaigatqiaiglaigqsvvcs 236  
 QY 281 DNDVNAANVR-AYAPDNDQPIDNRKATFKNGATDVFSGNSGNDISIRKIIINWAGSA 339  
 Db 237 dn-----nsrpaytp-ntgaldpkfgatmtkagpl-----sigsnsikrklinvagayn 285  
 QY 340 DTDVAVNAQLEAVRLA--NQITFEKDDSNRVEKGLGTLITTCAGQTSALTDRHIGVY 398  
 Db 286 ktdavnyaqgleavvkwakerlftfgddnsatdvkigldntllikygaaetnalndniygv 345  
 QY 399 QNGD--GLKQLAETLTLKLVNTENTLANKVTYVK----- 433  
 Db 346 keadnsqkvlakltnlnlevntltlnaltlvkvssstetlellsdlftqpnbsq 405  
 QY 434 -----TRLTPDKIGFTNDMNGIDESKPYLDK----- 459  
 Db 406 stektvygvnvgvftmaettaigttrtrdkigfaragdvekekqpyldkqkqkvgsv 465  
 QY 460 ---DTGIIHAGGQKRTKLVAGVDDDAATYQDK----- 489  
 Db 466 altidngidagunklslakgssandavtleglkaekpnlinaagajsviprteisvskag 525  
 QY 490 -----KVNQTAESALQTF 502  
 Db 526 nvtapryniqvkktelnsdgtscdfsvksgtmsjlvtaehlsylnvnrtdsalsaf 585  
 QY 503 TVKKVDKNGDANDSKITITVGKNNKPDGTQVNTLKLKGENGVDVTEETNGTFEGNUNN 562  
 Db 586 tvke-----edddanaetvakotdknagavslklkgknglvackkdgvtvfglsqgs 640  
 QY 563 GLTVGNSLTNNDGLSVKNTNSNKOIOVGADITFTDISNSKPGAGIENTRTRIDGIGRA 622  
 Db 641 glitgkstlndngltvtkd--neqiygavngikftlvngsnpqglntlanlarlt----- 692  
 QY 623 NNTGSLDANKRLPTPTGIMCGKELTVQSAINPATNGQLDPMNRNSTANTEKSSAAT 662  
 Db 693 ----- 692  
 QY 663 IKDLYNLSQVLPFRAGDTPGNVTKLGEILKVGKTTADLTKNIGVAVDSTNSLTJV 742  
 Db 693 ----- 692  
 QY 743 KLAFTSLDLAVNKTTLTASDKVTVDGNNATKLQNGDLTFPSKONTGATPATNSKITIGVD 802  
 Db 693 ----- 692  
 QY 603 GLKFTDNNGIALDGTYYITTKKVGFAKQDGLSKRPYLDKLVKGEVETTINGINAG 862  
 Db 693 -----dkigfagsdgaavdnkpyldgdklgyvnnvltntglnag 732  
 QY 863 KATIGLSNLTLDATNATGHTVQGIYVSDTKTPRAISIGDVLANAGFNLKNNGDAKDFVST 922  
 Db 733 kaigtlsptlpsiadgsrnlgnltgdkksnaasindlntgfnlknmpidfvst 792  
 QY 923 YDVFDFINGNAFTAKVYD--GKASKVAAYDVNVDTGIIHLTGADGNKNGQIGVTTITKT 980  
 Db 852 saangtlnhtvnsdadaivnakdaenlnclakehlttqgtadtaqlctfvkqvdeem 911  
 QY 961 DAKGKATNFVNSGDDKALINAKDIADNLTLAGELIRNRTGTPALQTFQFQYKVEENG 1040  
 Db 1041 DDDNDADTITVIGDAKTQNVVTLKLKGNGIDQTFNKGDFVTEGINTQSGIKAGNNTTN 1100  
 QY 912 nad--danaetvvgqnamgnvntltlkengnlnktkdnngvtvftintsglskag--kstln 969



```

Db 237 dn-----nsrpaycp-ntqaldpkrfgatnmtkagpl-----slgsmskklkinnvgavn 285
QY 340 DTDVAVNAOLKEAVRLA-NRQITFKGDDSNRRVEKGLGKLTITTGAGQTSALDDHNGV 338
Db 286 ktdavnaqgleavvkwakeritltfgddnstdvqigldntllkysgaetnaldnngv 345
QY 399 QNGD--GLKQJLAETTLTLKAVTTEENTLANEKVVGK----- 433
Db 346 keadnsglkvklaktlnlnevtnttlnatltkvvgssstaeallsdaltftqpnbsq 405
QY 434 -----RLLTDTKIGFTNDNGIDESKPYLDK----- 459
Db 406 stsktyvgvngvfktnaetaaigttrirckifgarfdgvdckqpydkkqklygsv 465
QY 460 ----DTGIIHAGGOKITKLTAGVDDDAATYGOLK----- 489
Db 466 altldngldsgnkkisnlakysandavtlleglkaakprlnaagajlsvprtleisvdaksq 525
QY 490 -----KVNQZAESALQTF 502
Db 526 nvtaptyngvkttelnsdgtaskfayksgqtnslvtaehlasylnevnrtadsalqst 585
QY 503 TVKKVVKNGNDANDSKITTYGKNNKPDGTOVNTLKLKGENGVDTTETNGTVEFGINQNN 562
Db 586 tvke-----edddanaltvakdctknagavslklklygnlvtackdgtvftvglsqds 640
QY 563 GLTVNGSTLNNDGLSYKNTNSNKOIOVAGDITFTDISNKGAGIENRTTRIDRIGRA 622
Db 641 gltfgstlndgltykdc--neqiygagnglftfnvngsnpgtylanarlttr----- 692
QY 623 NNTGSLDANKPRLPTGINAGKELTNVQSAINPATNGQLDPMRLSTANTEKSSAAT 682
Db 693 ----- 692
QY 683 IKDLNLSQVPLTFAGDTGPNVTKKIGELIKVAGGKTADDLTKNNIGVADSTDSNLTIV 742
Db 693 ----- 692
QY 743 KLAKTISLDLAVNTKLTLASDKVTVDSGNNTAKLQNGDLTFESQONTGATPATNSKITGV 802
Db 693 ----- 692
QY 803 GLKFTDNNGLADGTTYITRKDVYKFAKODGSLDKSPYLDKDKLKVEYEITNGINAG 862
Db 693 -----dkigfagsdgaavdnkpyldqklgygnvklntngnag 732
QY 863 KATGLSNTLTDAATNATGHWVQLGIVDSTDKTRASIGDVLNAGPNLKNNDADAFVST 922
Db 733 kaltglspclpsladqssnrlngtltqdkksnaasindlntgfnknmnpidftvst 792
QY 923 YDTVDFINGNATTAKVTYD--GKASKVADVAVNDGTTIHLTGADGKKNQIGVKTTLTFT 980
Db 793 ydtvdfangnattactvhtltanktsrvydvavndtclhlgtddnk-klyvktclinkt 851
QY 981 DAKGKAIFSVNSGDDKALINAKDIADNLNLAGEIRNTKGTADTALOTFOVKRYKENG 1040
Db 852 saanglatnfnvnsdedalnvakdiaenlntlakeihltctadtaqlftfvkvvden 911
QY 1041 DDNDNDADTTVKGDAKTNQVNTLKLKGKGLDIONKOSTVFEINGTOSGLKAGNNTILN 1100
Db 912 nad-danaaltvgknaannvntclikgenglnikctdngltvfigntstgkag-ksctln 969
QY 1101 NNGLSIKNTAGNEOIOVAGDGVFAKV--NGVYAGAGIDGTTITRDEIGFAGTNGSLDKS 1159
Db 970 dsglslknptgseqiygvsdgvkfakvnmngvvgagidgtttrttdelgftgtnslks 1029
QY 1160 KPHLSKDGINAGGKKTITNIOSEIQAONSNDVAVGKIYDLKTELENKISSTAKTAQNSLH 1219
Db 1030 kphlskdginagggkkltnqsgelagnsdavtggklydlkteleenklisstaktagnslh 1089
QY 1220 EFSVADDECGNNTFVSNPYSSYDTSPTVTFAGENGITTVYKKNKVAVNGIDOTGLTTP 1279
Db 1220 EFSVADDECGNNTFVSNPYSSYDTSPTVTFAGENGITTVYKKNKVAVNGIDOTGLTTP 1279

```

```

Db 1090 efsvadeggnftvsnpyssydtstsdvltfagenglttkvknkvrvvgldgtqgltp 1149
QY 1280 KLTVGNNGNGJLVIDSQONCTITGLSNTLANVNTDKGSRVTTGOKIIRKEDKTRASI 1339
Db 1150 klvgnngkgjlvldsgnqntltglsnltanvntdnkgsrvrtlegnllkdedktraasi 1209
QY 1340 VDLVLSAGFNLQNGEAVDVSSTVDVYVFPADGNAATYATYDTSKTSKVYVDVNDVDTTI 1399
Db 1210 vdlvlsagfnlqngaeavdvslydvtvnltadgnattakvtyddtstkskvvydvndvdtclt 1269
QY 1400 EVKDKKLGVKTTTLTSTGANKFALSNOATGDLVYKASDIYAHNTLTSQDLOTAKGASQ 1459
Db 1270 evkdkklgvtltltstgkankfalsnagtdalvksadlvahntlsqdlgtakgasq 1329
QY 1460 ANSSAGYVDADGNKYIYDSTDNKKYIOAKNDGTVDKTEKVAADKILVAAQOTPDGTLAQMNV 1519
Db 1330 anssagvydadgnkylvstdnkkyygakndgtvdktevakdklvaaqrpddgtlaqmrv 1389
QY 1520 KSVINKBOYNDANKKOGIENEDNAFYKGLEKASDKKTKNAAVTVGDLMNAVQOTPLTFAGD 1579
Db 1350 ksvinkboyndankkgylnednalfvglekaasdnktnaaavtvgdlnavaqrpdlftag 1449
QY 1560 TGTAKKLGAEETLTIGQOTDKTLTDNNIGVAGTGTFTVKLAKDLTNLSNVNNGTKID 1639
Db 1450 tgtakklgeltltlkgqtdntkldnmlgyvaagdgftvklakdltnlnsvnngtkid 1509
QY 1640 EKGVSFVBSGAKANTPYLSANGLDLGGKVIYSNVKGCTKDTDAANOQLEVERNLLGLG 1699
Db 1510 ekgvsfvsbgakantpylsangldlggkvisnvkgctkdtdaanvqglevernllglg 1569
QY 1700 NAGNDNAGQONVNIADIKKDPKSGSSNRVTYIKAGTVYGVGKGNNDTEKLANGVQVGVND 1759
Db 1570 nagndnagqonvniadikkdpksgssnrvtlykagtvlygvgmndteklacvlygvvdk 1629
QY 1760 DGNANGDLSNVVWKTQKDGSKRALLATYNAAGQTNVLTNNPAAEADIRINEGIRFFHVND 1819
Db 1630 dganangdlsnvvwtqkdsgskrallatynaagqtnvltmpaaeidrlnegriffvnd 1689
QY 1820 GNOEPVVGKRGNTSSSAGKSHVAIGPOAKADGEAANAIGQTOAGQNSIAGDMAATG 1879
Db 1690 gnoepvvgkrgntssagkshvaigpoakadgeaanaigqtoagqnsiagdmatg 1749
QY 1880 DOSIATIGTNVAVAGKHSAGIADPSTVKADNSYSGVNNNOFIDATQOTVFGVNNITVTES 1939
Db 1750 dgsiatigtvavagkhsagiadpstvkadnsysgvnnnqfidatqotvfgvnnitvtes 1809
QY 1940 NSVALGNSAISAGTHACTOAKKSDGTAGTTTATGATGVYKGFAGQTAAGVAVSGASGAE 1999
Db 1810 nsvalgnsaisagthactoakkssdgtagttttagatgvkykfgagtaavsgasgae 1869
QY 2000 RRIONVAAGEVSATSTAVNSOULYKATOGIANTTNLDRIRIHENKANAAGISSAMAMA 2059
Db 1870 rriionvaagevsatstavnsoulykatogianttnldririhenekanaagissamama 1929
QY 2060 SMPQAIYFGRSMVGIATIHNGQAVAVVAGLSKISLSDNQCMFKINGSADTQGHGAAGVAG 2119
Db 1930 smpqaiyfgzrmvvgiatihngqavavvagsklsldngqvwfkinsadtcghvgaavag 1989
QY 2120 FHF 2122
Db 1990 fhf 1992

```

RESULT 6  
 AAB69137 standard; Protein; 1992 AA.  
 ID AAB69137  
 AC AAB69137;  
 XX  
 XX  
 DT 24-Apr-2001 (first entry)  
 XX  
 DE M. catarrhalis M56 200kDa protein in PKS348 SEQ ID NO:13.  
 XX

[illegible]

QY	399	QNGD--GLKYOIAEITLILSKMVTTEMLPANENVYVGK-----	433
Db	346	keadnsglkyklaktlnlnltevntllnatltvkwsssstlaelJdsJltfcpntsg	405
QY	434	-----TRLTPTKIGFTMDMNGIDESKPYLDK-----	459
Db	466	stsktygvngvkvfcmnaetlaigttrlttrclgJfardgvdekgppJldkqglkvgsV	465
QY	460	-----DWGIIHAGGOKIRKLTPTAGVVDDEDAATVGOLK-----	489
Db	466	aJltldngJdggnkklngJlkgssandavJleqJlkaakpJtlagagJsvJtpeJlsvdaksG	525
QY	490	-----KVNQTAESAIAQTF-----	502
Db	526	nvtaPtyngJyktktelnsdtsqkfsvksgJfmsJlvaehlaJsyJlnevtmdaJlsJf	565
QY	503	JYKVKVDKNGDANDSKIIYNGKRNKRDGQVNTLKLKGBNGVDVTEFNGTVFGLNQNN	562
Db	566	lvke-----edddanaJlvaKdJtlnaagavJlJlkgngJlvaatkdgJvtJlgJsgds	640
QY	563	GLTVGNSTLNNDGJLSTVYNTSNKNOIOVGADGJTFPFDISNSKPGAGJENTTRJTRDGI	622
Db	641	glJtfgskJtlnmdgJtvtcd--neqJvgngngJlfnvngsmpgJtJantJrJt-----	692
QY	623	NNTGSLDANKPRLTPGJNNGKELTNVOSAINPATNGQJDFMRLSTATEKSSGAAT	662
Db	693	-----	692
QY	683	JkDLYNLSQVPLTFAGDPTGNNTKJLGJELLKVGGKTTADJLTKNNGVADSDNSJLV	742
Db	693	-----	692
QY	743	KLAKTJLSDLDVNTKJLTJASDKVTVDSGNNTAKJONGJLTFESKONTGATPATNSKJTG	802
Db	693	-----	692
QY	803	GLKJTFDNNGJALDGTJYTKDKRGPAKODGSLDKSKPYJDKDKJLVGEVJETTNGJNAG	862
Db	693	-----dkJgJsgdgJdvJtnPpyJldqkJlvgnvkJltnJcJlagnag	732
QY	863	KJITGJLSNTJLDATNATGTHYQJGVJDSYDSDKJRAASJIDVYNAGJFNJKNNGDAKJ	922
Db	733	keJtJgJspJtJpsJldgJdsrJnJelgJntJqdkJdksnaasJndJntgJfnJkmmnpJldf	792
QY	923	YDTPVFIINGNATTAFAKJYVD--GKASKVAVDVNVDTJTJHLTGADGNKNOJGVKTTJLKT	980
Db	793	JyJlJvfJangJntJctJvJhdJankJskJvyvJnvJdJtJlJtJgdJdnK--kJgvkJtJkntK	851
QY	981	DAKGDKAJNFVSNGDDKALJNKAJDIADJNJTJAGJRNTRKGTATJAJQTFQVKKVYENG	1040
Db	852	sangJtJatJnJfJvnsJdJedaJlvaJnJdaJlJaeJnJtJlJakeJlJhtJktJadJtaJqJtJv	911
QY	1041	DDDNADJITTYGJKAJKTNOVJTLKJGKNJLDJQTKKDGJVTFGJINTGSGJLAKJNNJTLN	1100
Db	912	aad--danajlvJgJknaJngJvtJlJkxgJngJmJktJkgJvJlJfJntJsgJlJkag--ksJlJn	969
QY	1101	NNGLSJKTJYAGNEOJYOVGADGVKFAKY--NNGVYGAGJIDGTJTRJTRJDEJGFAJNTGSL	1159
Db	970	JggJlJstJknpJtJsgJqJvgJadgJvtJkTJrJnmJngJvJagJldgJtJrJtJdeJlJfgJtJng	1029
QY	1160	KPHLSKDJGJNAGGKJTTJNIGSGEJLAONSJNDVJTGKJYJLDKJLEJLEKJISSTAKTJQNSJLH	1219
Db	1030	kpJlJskJdgJlJnagJkJlJtnJlJgsJlJagJshJdJvJgJkJyJldJkJtelJenJksJstJk	1089
QY	1220	ESSVADDEGNNFTVNPJSSYJDSJTSKSDVJTFPFGENGJTTKWKJGVJRVJGJDOJTGJLTP	1279
Db	1090	elsJvJadeJgJnmJtJvsJnpJssJdsJctJsdJvJlJfJgJngJlJtJknJkgJvJrvJgJldgJtJ	1149
QY	1280	KJLVGNNGNGKJIVJDSQNGJTTJIGJLSNTJANTJANDGJSVJRTJEGKJIKJDEKJTRAASI	1339
Db	1150	kJtJvgJnmJngJkJlJvJdsJngJntJlJcJlJnJlJantJndJkgsJvJrtJlegJnJlJk	1209
QY	1340	VDVJLSAGJNVLGJNGEJAVJFSJTYJTVJFADGNAKTJAKJYJYDJSKJSKJVVYVNVNDJOTJ	1399

```

Db 1210 vdlvlsagfllqngneavdvstydvtvfnfagdnatleakvdyddtsktskvvyvnnvddtli 1269
QY 1400 EYVDDKLLGYKTTTLTSTGTGANKFALSNQATGDALVKAADYIAHLNTLSGDIOTAKGASQ 1459
Db 1270 eyvdkklyvktctllscgcygankfalsngatgdalvkaadlvahnlntlsqddqktakgaag 1329
QY 1460 ANSSAGYVDADGNKVIYDSTDNKYVQAKNDGVDTKEVAKRKLVAQAQPTPGTTAQMNV 1519
Db 1330 annsagyvdadgnkvlystdtnkkygkndgvtvdktkerakklvaagqtpgltlaqmny 1389
QY 1520 KSYINKEQVNDANKKGGIENEDNAFVKGLEKASDNKTKAAATVGDNLNAVQPTLTFACD 1579
Db 1390 ksyinkeqvndankkgylnednaftvglekaadnktknaavtygdlnaavqptltfagd 1449
QY 1580 TGTATKLLGETLTIKGGOTDNKLTDDNNTGVVAGTDFGVKLAKDLTNNNSVAGGTRKD 1639
Db 1450 tgtatklletltikggqddctnkltdnnlgyvvaigdfvklakdltnnsvnaagqtkld 1509
QY 1640 DKGVSFVDSGQAKANTPVLANSGLDLGKVIISNMGKGTDPDAAVQOLNEVRNLGLG 1699
Db 1510 dkgsfvdsqgakantpvlansgldlgkvlisnvgkgtdddaanvqdlnevrnlglg 1569
QY 1700 NMGNDADNQNINADIKKDPNMGSSSNTVTKAGTVLGGKGNNTDKLATEGVGVGDK 1759
Db 1570 nmgndadngvniadikkdpnmgsssnrtvltkagtlvgkgnndeklatgvgvgydk 1629
QY 1760 DGNANGLDLSNVVWTKQKDSKKALLATYNAAGOTVLTNNPAEALIDRIIEGIRFEHND 1819
Db 1630 dgnangldlsnvvwtkqkdgskkallaltynaagotvltlnnpaeadrlieqgtrfthvnd 1689
QY 1820 GNOEPVVGKRGIDSSASGKSHVAIFQAKADGEAAVAGROTOAGNOSTAIGDNAQATG 1879
Db 1690 gnoepvvggrngidssasgkshvaifqakadgeaavaigrtgqngsialgdnagctg 1749
QY 1880 DOSIATGCGNVVAGKHSGLIGPSTVYKADNSYSGVGNNOFTPATGTDFEGVGNNTTVES 1939
Db 1750 dosiatgcnvvagkhsaglpgdstvckadnsysvgnnftatldtfdygvgnnlvtles 1809
QY 1940 NSVALGSNSAISAGTHAGTOAKKSDGTATTTAGATGTGVKGFAGOTFVAGVAGSAGAE 1999
Db 1810 nsvalgsnsaisagthagtoakksdgtatgtagtltvgyfagtagvavagsagae 1869
QY 2000 RRIQNVAAEGVASTDAVNSQLYKATQOGANATNELDRIHONENKRNAMGISSAMAMA 2059
Db 1870 rriqnvaaegvastsldavnsqlykatqsiatnecneldhrilqnenkanagissamama 1929
QY 2060 SMPQAYIPGRSMWGTGIAITHNGOGAVAVGLSKLSDNGQWFKINSADTQGVGAAGAG 2119
Db 1930 smpgaylpgirsmvgtgiatthngogavaavglkslscngqwklingssadtqgvgaagag 1989
QY 2120 FHF 2122
Db 1990 fhf 1992

```

```

XX 01-FEB-2001.
PD 26-JUL-2000; 2000MO-CA00870.
XX 27-JUL-1999; 99US-0361619.
PR (CONN) CONNAUGHT LAB LTD.
XX Logjampro SM, Sasaki K, Yang Y, Klein MH.
XX WPI; 2001-159722/16.
DR N-PSDB; AAF59105.
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
PS Claim 1; Fig 5A-Y; 247P; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis les1 200kDa protein, which is given in the exemplification
CC of the present invention.
XX
SQ Sequence 2314 AA.

```

Query Match 61.6%; Score 6600; DB 22; Length 2314;  
 Best Local Similarity 57.1%; Pred. No. 7; le-274;  
 Matches 1460; Conservative 133; Mismatches 278; Indels 682; Gaps 39;

```

QY 1 MNIHYKVFNKATGTGMAVAEYAKSHSTGGSCATGCGSVRTLSFARIALAVLIGAT 60
Db 1 mnhlykvfnkatgtgmavaeayakshstggscatgcvsvrtlsfariavllvlgat 60
QY 61 LNSGAYVAGIGISEMDGK---GANARGDSIAIGDAQALGOSTAIGNKK-IVNNSNN 116
Db 61 lngsavyagigisemdkg---ganargdsiaigdaqalgsostaignnkv-ivnnsnn 116
QY 117 NANIGAKASGNESIAIGDVLASGHSIAIGSDLYLKRTVYO---ISELLEPIINGOKA 173
Db 117 nanigakasnesiaigdvlasghsiaigsdlylkrtvyo---isellepilingoka 173
QY 121 ngkgsnaksksesialsgdvlaegdasialgsddlydrnstskypngllstlqnhv 180
Db 121 ngkgsnaksksesialsgdvlaegdasialgsddlydrnstskypngllstlqnhv 180
QY 174 LNDIYQIADTN-LQKVRTAAGHASTAVGAMSYANGHFSNAGTAADEGYSLAVLT 232
Db 174 lndiyqiadtn-lqkvrtaaghasavagamsyanghfsnagttaadegyslavlt 232
QY 181 lr---qirdngsqkyrrlaeghaslavgamayakghafnagtrsteagmslavlt 237
Db 181 lr---qirdngsqkyrrlaeghaslavgamayakghafnagtrsteagmslavlt 237
QY 233 ATAKAASSIAVGSNAOIGAFAATVAVGSGTOVNLNRGIALGFGSOVLOKNDVAAAVRAY 292
Db 233 atakaassiavgssnaoigafaatvavgshtovnlnrqialgfgsovloknvdaavray 292
QY 293 APDDNOPIDNRKATFENGATDVESIGNSGNDISIRKTIINVGAGSADTDAVVAOLKEA 352
Db 293 apddnopidnrkatfengatdvessignsgndisirktiinvgagsadtdavvaolkea 352
QY 295 vpegnsgnlskskat-gng---lfsigs-----tlkrklinyaggyedcdavnavglkev 346
Db 295 vpegnsgnlskskat-gng---lfsigs-----tlkrklinyaggyedcdavnavglkev 346
QY 353 VRLANRQITFKGDSNNRVEKGLKTLTTGG-AQTSALTD-HNIGV-ONGDGKLVOLA 409
Db 353 vrlanrqitfkgsnnrvekglkltttgg-aqtsaltd-hnigv-ongdgklvola 409
QY 347 enlakrgitlkgdngltvkkigetltikgsgetakldtlnnlgvcdantlgtkvla 406
Db 347 enlakrgitlkgdngltvkkigetltikgsgetakldtlnnlgvcdantlgtkvla 406
QY 410 EFTLSIKMVTENTANNEKTVGK----- 433
Db 407 knlsgletvstknltasekvtsngmtaelsgglltftptnascdkcvygtgdlkftd 466

```

QY 434 -----TRLTDTKIGFTNDMNGIDESKPYLDKDGIIHGOKITKLAGVDDDAAT 484  
Db 467 nsntaledttrtkdkgfsnkgatvenkpylkd----- 502  
QY 485 YGOLKVVNQTAEBSALQFTTVAKKVDKNGDANDSKIIIVGKNNKPDGTQVNTLKLKGENV 544  
Db 503 -----klk----- 505  
QY 545 DYTETNGTVPFGLNONGLTGVNSTLNDGLSYKNT--NSNKIOVGADGITTDTDISNS 602  
Db 506 -----vgnstlmgglvntnlgsgnkqiyvsgadikladavn 544  
QY 603 KPGAGIENTRITRDGIGFANNSTSLDANKPRL-----DPTGINAGKELTN 649  
Db 545 vsmnagfgttrteeieygfadaqkvdkspylkqkqiygvvktlctdsqimnagdkksh 604  
QY 650 VOSA-----INPATNGGOLDFMNRRLSTANTEKSGSAATIKDLVNLISQVP-- 693  
Db 605 vkdatdtdavtyqklqvsgadqalsfslr-----dekqgefslmlysnqntpnt 658  
QY 694 ---LTFAGDTPNVTKLIGELIKVKGKTTADDLTKNNIGVADSTDSLTVLAKATLSD 750  
Db 659 fclttagengisi-----sndaikgkvkv-----g 684  
QY 751 LDAVNTKTLTASDKVTVDSGNNTAKLNGDLTFSKONTGATPATNSKTIGVDGLKFTDNN 810  
Db 685 idpnl-----gltpkl-----lvsgd----- 701  
QY 811 GIALDGTYYITRKDVGFARODSLSKSPYLDKDLKVEEITTINGINAGKATIGLSN 870  
Db 702 ---xdkgtqlvlegvasndklnl-----lrglsp 728  
QY 871 TLTPATNATTHVYQLG--IYDSTDKTRASTGCVYLAGFNKNGCAKDFSVTYDVPF 928  
Db 729 clpsltnagvrtlegnltlsdckskasigqilntgflknmsvqfvsylntvdf 788  
QY 929 INGNATTAKEYD--GKASKVAYDVNVDGTTILTGADGNKNOIGVTTTLTKTKGADGK 986  
Db 789 idgnatekavvdebtngskvtyvndektietlgnngkunklgyvktlctlttnang-k 847  
QY 987 AINESVNSGDDKALINAKDINLNTLAGELRNTKGTADALQTFQVKKYKENGDDND 1046  
Db 848 atnslst--dnalvnaeklaenlntlakelhttkgacalqtfkvk-----dgsld 900  
QY 1047 DTRVVGDAKTN--QVNTLKLKNGLDIOTNKDVTVPFGINTSGLKAGNNTLNNGL 1104  
Db 901 etltvgqdgltngkltvntliklgenqiltvatnkdgtvltfgntsglkgadstlknqgl 960  
QY 1105 SIKNTAGNEOIOVGADGVKFAKVNNGVYVAGIDGTTTRITRDEIGFAGTNGSLDKRPHLS 1164  
Db 961 slkmpasneqiyvsgadqvtakvdkgnstgldtstlctdqlygtgansldtkphlt 1020  
QY 1165 KD-----GINAGKKITINIOSGETAIONSNDAVTGKLYDKTETEENKISSSTAK 1212  
Db 1021 kdklkygeveltnglnagvkkltngsgdltngsndavagrvydlktelesksinaak 1080  
QY 1213 TAQNSLHEFVADBOGNNTVSNPYSSYDTSKTSIDVTTFAGENGITTKVNGGVVAVGIDQ 1272  
Db 1081 tagnsllhefsvadeqghlftvsnpyssydsctsktsdvltfagenglttkvngvrvyidq 1140  
QY 1273 TKGLTTPKLTGVNNGNGGIYDSONGNTITGTSNLTANTYNNKGSVRTTEOGKIINDED 1332  
Db 1141 tkgltpklvgnngngkxiyidskdgqntltgtsnltanlynd-gaghalasq-landtd 1198  
QY 1333 KTRASTIVDLASGFTNLGNGEAVDFVSTYDVTAFADGNATTAATVDDDTSKTSKVYDV 1392  
Db 1199 ktrastivdlvlnagfnglgngeavdvfsydvtfidgnattakvtyddcsktskvvdyv 1258  
QY 1393 NVDDTTIEV-KDKKLGKVTYTTTSTGTGAN--KFALSNQATGALVAKSDIVAHNLTLSC 1449  
Db 1259 nvdkltlevsckkllyvktlctktsangnackf---saadgdlvksadiathlnltag 1315

QY 1450 DIOTAKGASQANSAGYVDADGNKRYIYDSTDNKYQAKNDGTVDKTEYAKDKLVAQAOT 1509  
Db 1316 dqtakgasassassayvadadgnkryydstdkryygnndkgyvdkkexvdkdlvagaqt 1375  
QY 1510 PDGTLAOMNVKSVINKEQVNDANKKOGIENEDNAFVKGLEKASADNKTAKNAVTVGDLNAV 1569  
Db 1375 pdgtlaqmnvksvinkevndankkgloednafikglennaekctktnaavtyvgldnav 1435  
QY 1570 AOTPLTFACDTGTAKKLGTLITIKGGQDTPNKLTDNNIGVAGTGFVYKLAKDLTJNLA 1629  
Db 1436 btptltfagdtgtakklgellcikgqctdnkltdnnlgyvagtgtfvciklakdltnln 1495  
QY 1630 SVNAGGKTKIDDKVSVFVDSGSOAKANTPVLSANGDL----- 1666  
Db 1496 svnaggttridekqistvdanagkancpvlasangldlgyktrlsnigaavddnnavnfkf 1555  
QY 1667 ----- 1666  
Db 1556 nevaktvnlmgnsnagaslpfvrtidangkplngtdgkpkqkikgagdkryyhananqrvp 1615  
QY 1667 ----- 1666  
Db 1616 dkdqkpltdaklanlaabgkpldaghyvaslgnsdaltlnlksltplqldtpntgna 1675  
QY 1667 ----- 1666  
Db 1676 usagqagslpslsaaqgsnaasvkvlnvgfnglqtnhngvdfvkaydtvntfngltgdlts 1735  
QY 1667 ----- 1666  
Db 1736 vrsadgtvmsntvntalaatdddnllikakdgkfykaddlmpngslakgksaasdktp 1795  
QY 1667 ----- 1666  
Db 1796 glslvmpnagksgtgdavalnlskavfkskdgtlttvsdgsisigdkdnssltskg 1855  
QY 1667 ---GKRVISNVGKGTDPDPAANYQOLNEVRNLLGICLNACNDNANDNQVIAIDIKKPNNG 1723  
Db 1656 lnvgykvlsvngyktcdtdaanvgqlnevrnllgylgnagndagdnqvnlaiddkkpnsq 1915  
QY 1724 SSSNRVTIKAGTVLGGKGNNDTEKLTAGVOYGVNDKDGANCDLSNWWKTKODGSKKAL 1783  
Db 1916 sssnrvtikagtvlggkgnndteklatgvyqvgvdkgangdlsvnwvktkqdksgskal 1975  
QY 1784 LATVNAAGQTYLTNNPAEADIRINEOGIRFPHVNDGNOEPVVOGANGIDSSASGRHVA 1843  
Db 1976 latvnaagqtnyltnnpaealdrineqgtrfthvndgngpvrngngidsasgfhva 2035  
QY 1844 IGRQAKADGFAAVALIGROVQAGNOSIAIGDNQATGDSIAIGTGNVVAGKHSAGIDPS 1903  
Db 2036 lgrqakadgfaavaigrqtaqngsialgdnagatgdsialgtgnvvvgkhsagaldps 2095  
QY 1904 TVRADNSYSYGNNOPTDPAOTQDFVGVGNNTVYTESNSVALSNSAISAQTHAGTOAKKS 1963  
Db 2096 tvradnsysygnnngtldatqctvfygmnltvteensvalnselsagltagqaks 2155  
QY 1964 DGTAGTTTATGATGVKGFAGGTAVGAVSVGASGABERRIQNVAAAGVSATSTDAVNGSQL 2023  
Db 2156 dgtagttttagatgvtkglagtaglavavsagasgaerrlqnvaaagvsatstcdavngsq 2215  
QY 2024 YKATOGIANTNEDLRIRIQNEKNAAGISSAMAMASMOQATIPGSMYTGIGLAIHNGG 2083  
Db 2216 ykatgianaetelhrirhnenkanaagissamamasmpqaylprsmvtyglaiatnng 2275  
QY 2084 AVAVGLSKLSDNGOWFKXINGSADTQGHGAAVAGAFHF 2122  
Db 2276 avavglsklsdngowfklngsadtqghvgaavagfhf 2314  
RESULT 8  
AAR9393  
ID AAR9393 standard; Protein; 2353 AA.  
XX

[illegible]

Dh	231	-----esvdlvasgmnevlitgdknllidvltakengtlievtkpts-----vilek	279
Oy	402	DGLKVQAELEFLTSLKWTTEENLANKKVVNGKTRRLTTOXIGSTNNMNDISDKPYLDKT	461
Dh	280	dg-----klftgkennndtnkvlt-----sntadn-----tdegmvlvtrakavid---	318
Oy	462	GHAGGOKITKPLAGVDDDAATYGGOLKRVNOTAESALOTPTFKKVDKNGN-----DA	514
Dh	319	avnkagrvvtlttangngpfatva--sgtnvtfessgdgttasvctkdnngnglvtvkdyak	376
Oy	515	-----NSKILTYGKNNKPCGTQVNL-----KLKGENVDVT-----	548
Dh	377	vvgdglkfdsdckklvadttaltvtlvgkvaaelakeddckklvnaqdlvtalgnlswakaea	436
Oy	549	ETNCTVFETGNNONNGGLTVGKSTLNNDGLSVK-----NTNSNKOIQVGAGITFTDISNSKP	604
Dh	437	dtvgale-glskdegvvaagctvffkagknulkvkqganltyslqdalgsletlsigtfn	495
Oy	605	GAGIENTTTRTTRDQIGF--ANNTGSLDANKPRLPTGTINAGKELTNVQSAINPATNGGQ	662
Dh	496	g-gndaktvlnkdgltltpgngngtclgtntlsiatvtdgklagnkakaltvasgdlr-ayddan	553
Oy	663	IDFNRRJST-----ANNE-----KSSSATINDVLNLSQVPLTFAG- 698	
Dh	554	fdvlnmsatdnrhvedayqgllnlnehkannkpylvtscactvaygdllklygwvstkngt	613
Oy	699	DTGPNVTRKUGELIKVKGKGTADDLTKNNIGVADSTNSLTPLAKTASDL-----D	752
Dh	614	keesngyqkgadevlfllgagatvtskseng-----khltlsvaeatkadcgylekgd	665
Oy	753	AVNFKTLTA--SDKYTVDSGKNNTAKLQNGDLTSSKONTGATPA-----	793
Dh	666	tlklkvngnubnlvtvngngtavevksg--tevkltgtdadrgkvltvakatandaqk	722
Oy	794	-----TNSKTIICVDGILKFTDN--NGILDDGTYITRKDYKGFAKOD	831
Dh	723	vavtkdvatainaaactvfkentltsidednprndgkdalagdtllfkgagnulkvkrd	782
Oy	832	G---SLDKSRPYLDK---DKLKVG-----EVEITTTNGIN-----AG	861
Dh	783	gknltfclaklevktakvadtlltignptlgttclatpknltstsdglnfaketaadsg	842
Oy	862	GKAI--TGLSNTLND--TNNATGTHVHVOLGIVSDSTKTRASIGDVILNAGFNILKNNDKAD	918
Dh	843	sknyvlygiatlltpepaagkshvdl--nvdatkksnaasleadvlrgwmngnngnvd	900
Oy	919	FVSTYTDVDT-----	929
Dh	901	yvatcydvtnftcdstglttvtvtqkadgyadvklgaktsvlkhngkllftgklldann	960
Oy	930	-----	929
Dh	961	gatlvseddqkdtgltvltaktvldavnksgvrtgsegataetgataenagnaetvts	1020
Oy	930	-----NGNATPKVYVYDCKASKAVAYDVNV--DGTTHLTGADGKNKOIGVKTTLTKTDK-	983
Dh	1021	vfnkngnatlatvskdhngnvnkvdyrvngdgkll-----gdckklvadttllvtlvgkv	1074
Oy	964	-----GDKAINEVNSGDKKALINAKDIADNLNLAEIR-----NRKGTADPLAQ--	1029
Dh	1075	svpagsanvn-----nnkklvnaeglatalnlsvtlakdyadgesegetdgevkgad	1128
Oy	1030	-----TFQV-----KKVEN-----	1039
Dh	1129	kvtflagknulkvkssekdfyslqdtlgtltsltlgttangrndgtvlnkdgltiltan	1188
Oy	1040	-----GDDNDADITTVYVKD-----AKTNOV	1060
Dh	1189	gaagatdasngntlsvckdgsagankeltvksaalktykdtqntadevgdkefhaavkna	1248
Oy	1061	NTLKIKGNGJDIOTNNKG-----TYTFEGINTOSGLKAGANNTLNNNG--LSIKTFAGNE	1113
Dh	1249	nevefvngagatvsaactdmngkhvnltdv-----aaakvvgdglekdcdgkrlklvndtngm	1305





Db 1 mklfnvlnwmtglwvsvseltthck--rasatvetavlallfatvganat----- 52  
QY 61 LINGSYAGIGISEADGCGGAGANARBDKSIAGDINO-----ALGSGSIAIDGNKIVHNSN 115  
Db 53 -----dedeqlavrtavrlsfhsdksgtgekevtensn 87  
QY 116 NNANIGAKASGNESIAIGDVLASGHASIAIGSDLLYKKEETVOQISELLPIINGOKALN 175  
Db 88 wglfthmg-----vfkagatlkag-dnlkikgkgt----- 117  
QY 176 DIYQLADTNLQKVRTHAOGHASTAVGAMSTAKGHFSNAFGTRATAEQTYISLANGLT-AT 234  
Db 118 -----nassf-----tyslkkdlt 134  
QY 235 AKASSIANGSNAOALGFAATVAG-----GSTOVNLNRGIALGFGSOVLKDNVDNMAN 288  
Db 135 svateklstfgangdkvdltsdanglklaktgngvhl-----gldscldpavrtcyvl 189  
QY 289 VAVAPDDNQPIDNRHYKATFKNGADVFESIG--NSNGNDSIRKRIIVGASADTDANVVA 347  
Db 190 ssfipnd-----vekratavk-----dvlmagvnlkg-----aktagnv- 226  
QY 348 QLKAVRL--ANRQTFEGKDSNSN-----RVEKGLCKTLITIGGAOTSAITDHNGVONG 401  
Db 227 -----esvdlvsaynveflgtcknldvvlakengktlevkftpkts-----vikck 275  
QY 402 DGLKQOLATLTLKMWVTENLTANEKVTVGKTRITTDKIGFTNMNGIDSKPYLKDIT 461  
Db 276 dg-----kltfgkenndtnkvt--sntatdn-----tdeagnjlvakavid-- 314  
QY 462 GIHAGOKITKLTAGVDDDAATYGOALKKVNQTAESALOTFTVKKVDKNGN-----DA- 514  
Db 315 avnkagwrvklttlangngdfatva--sgtnvfesgdtlasvckdltngngltkygk 372  
QY 515 -----NDSKIIT-----VGKNNKPDG-----TOVNTIKLGENGV 544  
Db 373 vgdgklfidsdkklvadltaltlvrgkvaekeddkkklmagdlvtalgnlswkaea 432  
QY 545 DUTTERNGTVTFGLNONNGLTGVNSLTNNNDGLSVK-----NTNSNKOIOGADGIFFTDIS 600  
Db 433 d--tdtdgale-glskdegvkgagetvftkagvnlkvkqdggnftslgdtlglstltg 489  
QY 601 NSKPGAGIENTRTIRFDIGF--ANNTGSLDANKPRLPTGIMAGKELTNVQASINPAT 658  
Db 490 gtlng--gndakvlnkdgltlcpagngtfgtntslvtckdgikagknaltvnaaglr-ay 547  
QY 659 NGQOLDPMRLST-----ANTE-----KSGSATIKDLNLSQVPLT 695  
Db 548 ddantfvlmsatdlnrhvedaykgljlnleknankqplvtidstaavgdllklgwwst 607  
QY 696 FAG-DTGPVWTKKLEILKVKGKTADDLTKNNIGVADVSDNSLTVLAKLTLD- 750  
Db 608 kngltkeesngvkgadevltltgagaatvtskeng-----khltsvvaetkdsagje 659  
QY 751 -----LDVAVNTKTLASDKVTVDSGNNTAKLQNGDLTFESKONTGATPA----- 793  
Db 660 kdgdtlklkvndqnt-----dnvltvgngtavlckg--fevtkgtdadrgkvtyk 710  
QY 794 -----TNSKTIQVDBLKFTDN-----NGIALDGTYYT 821  
Db 711 datandackkvatvkvataalnsaafvkteentltsidednpldngkddalagadltfk 770  
QY 822 KOKVGAPODQ--SLDKSKPYLDK-----DKIKVAG-----EVEITTINGIN 859  
Db 771 agknlkvrtidgknltfdlcknlekvktaksdltltlgntprtggtatcpkvnltstcadgln 830  
QY 860 -----AGGRAI--TGLSNTLTD--ATNATTGHVTLQIGVDSKTRASIGDVLINAGF 908  
Db 831 faketaasgsknylkglatiltltpesagakashval--nvdatkksnaasiedvltiragv 888  
QY 909 NLKNNGDADKDFSTYTVDFI----- 929  
Db 889 nlsgngnndvyaavltydtnlftdtslgtltvltvkgadgkagdvklgaktsvlkdngklf 948

QY 930 ----- 929  
Db 949 tgdklkanngatvseddydgtlgtvltaktvldavnskgrvltgagatavag 1008  
QY 930 -----NGNATTAKVITYDGKASKVAVDVNV--DGTIHLTLGADGNKNOIGVNT 974  
Db 1009 naeltsgtsvnfngnntaltvskdnglnvkvdydvngdglkl-----gddkklvadt 1062  
QY 975 TFLKRTDAK-----GDKAIFSVNSGDDKALINAKDIADLNLTLAGEIRNTKGTADTALO 1029  
Db 1063 tcltvvgkvavpaganv-----nkkllvnaeglatalnltstakadk-yadgeee 1115  
QY 1030 TFQVKKVKNNGDDNDADITTVGKDAKTNO-----VNTLKL-----KGKNG 1070  
Db 1116 getdgevka-----gdkvltfkagknlkvkqsekdfvyslqdtlgtlslgtlangrnd 1170  
QY 1071 LDIOTNRKG--TVTFGITVSGGLKAGNNTILN--NNGIS-----IKMT 1109  
Db 1171 tglvlnkdgtliltlangaaagtdasngntslsvckdgsagkneltvksalktykdqnt 1230  
QY 1110 AGNEQIQVGA-----DGVKFAKVNNGVAGIDG-----TTRIT----- 1143  
Db 1231 agatcpaanataevakqdlvdltkpatgaagngadakapdtlaatvgdlrglgrvlsakt 1290  
QY 1144 -----RDEIGFAGTNGS-----LDKSKPH-----LSKDG-- 1167  
Db 1291 adetqkfehaavknaneveftvgngatvsakctdnogkhlvtldvaeakvgdglektdg 1350  
QY 1168 -----INAGCKKITNIGSELIQONSNDVAVG-----GRIT-----DL 1199  
Db 1351 kklkvlndtgnlltvdatkgaavknefnavltdeitlagtuanenergkvkngat 1410  
QY 1200 KTELENKISSITAKTAONSILHFS--VADEQGNFTVSNPVS--SYDTSKTSVITF-AG 1253  
Db 1411 atetckkkvatvgdvakainaadaatfvvenddestlidsplddgandalagdtliltkg 1470  
QY 1254 ENGITTIVKNGV--VRVGIDQTKGLTTPKLVGNNGK-----GIVDSONGON 1300  
Db 1471 knlvkrdgnkntfalandlsvksatvsklsjltngknltscdkgljlnfakdsrtygd 1530  
QY 1301 T-----ITGLSNTLANVTDKGSVRTTEOGKIIKDDDKIRASIVDVLSAGPNLOG----- 1351  
Db 1531 anhlnglaastltdltlinsgat--tnjlgngltdeekkraasvkvlnaagvngvkvpa 1589  
QY 1352 --NGEAVDFVSTYDTPVNEFADGNATTAKVTD--DTSKTSKVVVDVNVDDPTIEVKDKKL 1407  
Db 1550 ngyenldfvatgtvtdtvsdkdltsvteskngkrtev-----kig 1633  
QY 1408 VKTTTLSTG--TGAN-KFALSNQAT-----GDALVKASDIIVAHNLNLSGDIOT 1453  
Db 1634 aktsvlkdngklftgkkelkannngvltetdgdkgenglvtaekavldavnkagrvkt 1693  
QY 1454 AKGASQAN-----SASGYVDADGNKVITYDSTPNKRYQAKKNG--TVDAKFEYAK----- 1500  
Db 1694 t-gangngndatavagstuvlfadngntlaevt-----kangslvkvrvkvadglkl 1746  
QY 1501 --DKLVAQA--QTPDGLAOMNVKSVINKBOVDANKKOGINEDNAFVGLKKAASDNK 1555  
Db 1747 dgdklvadtvlvtvadgvtvtapn-----ngdck-----fvdaasgladelnk 1788  
QY 1556 TKMAAVTVGDLNANAVQPLRTFAGDTGTTAKKLGELTITKGQDTFNKLTNNIGVAGTD 1615  
Db 1789 lswta-tagkegtgevdpanasag--evkagkvtfkag-----dnlkikgsgk 1834  
QY 1616 GFTYKLANDLNLNSV-----NAG-----GTRIDDKGVSPFSDSGQA-----KANTPYLSA 1661  
Db 1835 dftyslkelkltdlsvfekdangltgseestklitdgtliltipangagaaganantatvsk 1894  
QY 1662 NGILDGKGVISNVGKGT-----DTDAANYQOL--NEVRMLIGLGNAGNDN--A 1706  
Db 1895 dglsagknvntvngslklfkfgdhtlangvtdvdekhynadnykeltnldekgadnptva 1954



Db 585 aagaafaahyteatski-----tunvag-----nltagslda 616  
 Qy 658 TNGGOLDPMNRLSTANPEKSSSATIKDLYNLGVPILFADGPCPNPTKRLGELATKKG 717  
 Db 617 vngsqklctndnvtcnc-tlnactnltlnltda-vnlgjgdsllwnktagaafaahyt 673  
 Qy 718 KTTA--DDLTAKNNIGVAVDSTDNLSLFWLAKTLSDLDVAWTKPL-TASDKVYVDSGN-- 771  
 Db 674 dataeklnvtagnl-----tagstдавngsqklctndnvtcncnltat 716  
 Qy 772 NTAKLQNGDLTFESKQNTGAPPAINSKITGVGDLKFTDNGIALDGTYYITKDKVGAFAKOD 831  
 Db 717 nltitn--ltdavngjgdsllwnktaga--fssaahg--ltdateklnvtagadlta- 767  
 Qy 832 GSLDKSRPYLDKDLKXGEVEITNGINAGKATGTSNLTATNA-----TTG 881  
 Db 768 gstda---vngsqklctndnvtcnc-----tlnltln--ltdavngjgdsllwnktag 814  
 Qy 882 HVTOLGIVDSTDKTRAASIGDVLNAGFNLKNGDAKDFVSTYDVFINGNATTAATKYTD 941  
 Db 815 afaaahgtdateklnvtcncagad-ltag-----stдавng--sqklctnd 854  
 Qy 942 GKASKAVAYDVNWDGTTIHLTGADGNKNQIGVKTTLTK-TDANGDKAINFVNSGDDKAL 1000  
 Db 855 -----nvstnltlnltcdsvgdlk-----dssll 879  
 Qy 1001 INAKDIADNLTLAGETIRNTKGT-ADPALQTFQYKKVKEGDDNDMDITTVGKDKATNQ 1059  
 Db 880 -----wnkaagaafaahyteatsklnllagklnssnstdalngsqlygvadstcsy 930  
 Qy 1060 VNTLKLKGNGLDIQTKKDGTVTFGINTQSLKAGNNTTNNGLSIKTNAGNEOIOVGA 1119  
 Db 931 lg-----sgadl-----sdcgvlsgrplylg----- 951  
 Qy 1120 DGVAFAYVNNGVAVAGIDGTTTRITRDEIGFAGTNGSLDKSRPHSKDGINAGGKKTITNIQ 1179  
 Db 952 -gtclvngvdalaa lnstfsl-gdal lwdatagkfs-----akhnglnapsvltltda 1003  
 Qy 1180 SGELIAGNSNPAVNTGKITYDKTLEKNKISSPATAQON---SLHEFSYADQGNNTFYNSP 1236  
 Db 1004 ngavstscsdalngsqlygvadyadalggnavvntgslctpylaagssyn--vgda 1061  
 Qy 1237 YSSYDTSKTSQDVI---TFPAGENG-----ITPRVKNGVAVF---GIDOTKGL 1276  
 Db 1062 leaddtlldallwtdtangngafsaahgkdktaasvltvavgsatsndalngsqly 1121  
 Qy 1277 TTPKLVLYGNNGKGTIVDSONGONTTIGLSNTLANVT-NDKG-SVRTTEGOKTIKDEDKT 1334  
 Db 1122 stnkyiadaalvgdae--nadgltlapytiantdyinnvgaalldalldmalldwed-- 1175  
 Qy 1335 RAASIVDVLASGFPLAGNGEAVDF-----VSTYDYNFADGNMTTAKVYYDDTSKT 1385  
 Db 1176 -----agaynaashngnaskltnvaagdlstscdavnngsqlnatnllvt-----qn 1221  
 Qy 1386 SKVVYDV--NVDDPTIE-----VKDKKLGVTTLTSTGTGANKF--ALSNOATGD 1432  
 Db 1222 sqmnglnagntseyleengaglnyvrtnsgla findasssgjgataavynaavassav 1281  
 Qy 1433 ALYKASIVVAHLNLTSGDITPAKGAQANSAGYVADGKKVYYIDSTDNKYYAKKDGTV 1492  
 Db 1282 al--gqdsisevdt---glaigsssvss-----rvlwkgltn----- 1313  
 Qy 1493 DKTEVEVAKDKLVAGQAPDGTLAQMNKVSINKQOVNDANKKOGINDNMFVGLKKAAS 1552  
 Db 1314 -----tsvseegvnygdltldgel--lga lsl-----gddgkyrqln-----va 1351  
 Qy 1553 DNKTKNAVYVGD---NAVAOTPLTF--ACDPTGTAKKLETLTKGQOTPNKLTND 1606  
 Db 1352 dgsaahavavtrvqqlnaigavatlptkkyhanstaedslavgedslamgakt---lvg 1407  
 Qy 1607 NIGVAVGTGCTYVAKDLTNL-----NSVNAAGTAKIDDKG--VSFVDSGQAK 1653  
 Db 1408 nagjlgjln--tlvladalnglaigsnaranahadsiamngsgttrgagcnylaymndap 1465

Qy 1654 ANPVLASGLDLCGKVISNVGKTRPDPAANVOOLEVNRNLLGLGNAGNDAGNOYNI 1713  
 Db 1466 qnsvefsvybeegqrqlnvaagsadtdavvngq-----kvtld 1505  
 Qy 1714 ADIKDPNSGSSSNRTYVAGTVLGKGNNDPEKATGAGVGVVDKDNAGDLSNWWK 1773  
 Db 1506 aqvsqgtgtlnltngvt-----nldtr-----vnlengjgdlv----- 1540  
 Qy 1774 FQKDSKKALLATYNAAGOTNLTNNPAEALRINEOGIRFHVNDGQOEPPVVGCRNID 1833  
 Db 1541 -----tsgstkyfktn-----tdgad 1556  
 Qy 1834 SSASGKHSVAIGFOAKDEAVALIGRQAGNOSIALGDNQATGDOSIAGTGNVAG 1893  
 Db 1537 aaagkdsva l-----gsgslaaadn-----sva lgtsv-- 1586  
 Qy 1894 KHSALGDSPTVAKADSVSVGNNOFTDQTVDFGVGNINIVTESNSVA-LGSNSALISA 1952  
 Db 1587 -----adeentlsvgs-----tngrltcnvaagnatdavnvsqkxseaa--g 1628  
 Qy 1953 GHNAGTQAKKSPGCTAGTTTATGATGVKGFAGQTA VAVSAGSAGAEKRIQVAAAEVSA 2012  
 Db 1629 gyrvdtkadgsldysnltlggngsglt-----rlsvsag--v 1664  
 Qy 2013 TSTDAVNGSOLYKANGINATN---ELDHRIRHONENKANAGISSAMAMSPQAYIRG 2068  
 Db 1665 nmddavnvqqlkysvqetkqyldgrmvenmdnklskteslsjgslasamamqqlpqaylprg 1724  
 Qy 2069 RSWVTGCIATHNGOGAVAVAGLSKLDNCGQWPKINGSADTQGHVGAAGACGPHF 2122  
 Db 1725 asmasjggtylgesavalgsvmsangrvwyklgstinsggeysaalqagjqlw 1778

## RESULT 11

AAB23856.  
 ID AAR33856 standard; Protein; 1104 AA.

AC AAB23856;

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) protein from NTH1 strain 29.

KW Hia: adhesin; Haemophilus influenzae adhesin; NTH1: infection; vaccine;

KW non-typable; Haemophilus influenzae; antiinflammatory; auditory;

KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;

KW diagnosis; immunogenic; antigen.

OS Haemophilus influenzae.

PN WO200055191-A2.

PD 21-SEP-2000.

PF 16-MAR-2000; 2000WO-CA00289.

PR 16-MAR-1999; 99US-0268347.

PA (CONN:) CONNAUGHT LAB LTD.

PI Locmore SW, Yang Y, Klein MH;

DR WPT, 2000-618897/59.

DR N-PSDB; AAA92495.

PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for

PS use as antigens and vaccines and for treating Hemophilus influenzae

CC infection

Claim 1; Fig 20; 275bp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia)

CC protein from the non-typeable Haemophilus influenzae (NTHi) strain 29.  
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.

XX Sequence 1104 AA;

Query Match 9.5%; Score 1018.5; DB 21; Length 1104;  
 Best Local Similarity 27.8%; Pred. No. 9.1e-36;  
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;

QY 943 KASKVAVYVNDGTTHTLTCAGDGNKQIGVKTTLTKTAKGDKAIFNSGDD----- 997  
 DB 28 kcasatvaavaalatalatalsat-aeann-----tsvtnglnaygdtnfntnmsladelkv 81  
 QY 998 ----KALINAKD-----IADNINLTAGEIR-----NFKGTADTALOTFOYKKV 1037  
 DB 82 qdaykglhlnnekdnksflvadnlaetvgnlrlkglw'lskngrne--ksyqkqgd 139  
 QY 1038 ENDDDDNDADTIVGKDATNQNTLKLKGNKGLDIQTKKDDGVFNGICITQSGILKKNNT 1097  
 DB 140 e-----vltsgsaatvs-----ssskdghhtltsvltksfaevkt 176  
 QY 1098 TLNNGSLIKTAGNEQIOGADGVFAKVNNGVAGAGIDGTTTRTRDEIGFAGTNGSLD 1157  
 DB 177 dattg-----gynadrgrkvaeden--gadvd----- 202  
 QY 1158 KSKPHLSKDSINAGKKTINIGSGEIAONSNDVATGKTIYDLKTELENKISSAKTAQNS 1217  
 DB 203 -----kkvatvk--dvakaindaetfvkystoddliengaaqknettda 245  
 QY 1218 LHEFSVADEGNNFTV---SNPYSSYDTSKTSVITFAGENGITTVKNNGVAVGIDQTK 1274  
 DB 246 l-----kagdtcltlkagknlakldqngksyvtfalakldvtasakvsklsgkdnk 298  
 QY 1275 -----GLTTPKLTIVGNNNGKGIYIDSONGONTITGTSNTLIANTNDKGSVRTTEGCK 1326  
 DB 299 vdltsdanglkak--tgnng-----qngvnhlnglastl-----tdltlgtmtqaan 345  
 QY 1327 TIIDEDKTRAASIVDYLASGFNILOGNEAVPVSTYDVNFPADGNATTAATKAVYDDTSKTS 1386  
 DB 346 gvaevghnraasavadvlnagwlnqngasvdfvnaaydvfngntlnvnttdtakhkt 405  
 QY 1387 KVVYVD-----VAVDDTTIEVK-----IDKTKG-----VKTTLTSTGSG 1419  
 DB 406 tvrvdvtglrvpvgvtedgktvrvdnkyyeakqdsadmdkqvengelakltkvkvsasg 465  
 QY 1420 ANFEALSNQATDALYKASDIVA--HLNLTSGDIQTAKGASQANSAGYVADGNKVI-- 1475  
 DB 466 qnpvkltsnvaeg---teendavsfkqkalg-ekyvtltsasnayanngn--daggykatct 520  
 QY 1476 -----YSTDNKYYQAK--ND-----GTVDKTKVAVAKDLVAQAQTPDGTGLAQM 1517  
 DB 521 lnnqlnfkfstgdgellnkvendtvtlfpkksv--gvsgedgkatlqngklttdglveas 579  
 QY 1518 NVKSVIKKEQVNAKKGINENAFYKLEKASDNKTKNAAVTYGDINAQVQOTPLTFA 1577  
 DB 580 el-----veslnklgwkvvygdkgds-----geldgasnetlvks 613  
 QY 1578 GDTGTFAKKIGETLLTIGGOTDTRNKLTDNNIGVAVGTGDTFVKLANDLTMINSV----- 1631

DB 614 gdkvrl--kagenlkvk--qdgtn-----fyalakdelgvkvsvefda 654  
 QY 1632 ---NAGGKTKIDKGVSPFDSSG-----QAKANTPVLSANGLDGKVIYSNVGKTKPTDA 1683  
 DB 655 ngsnagskklkdgltltsangangaatadcklkvasdgsasgnkavknvsglkkfvgd 714  
 QY 1664 ANVOOLNEVRML-----GLGNMGNDADGNQVNIADIKRDPNSSSSSNTFTVAKG 1734  
 DB 715 anflptsadnlckpyddaykgltnlndekgadqkltvad-----nt-----aa 759  
 QY 1735 TVLGKGNNDIEKLATGSGVGVVDKDGANDLSNV--WTKTKGSKKALLATYNAAGQT 1793  
 DB 760 tv-----gdrlg'gwv--lsadkltgelnkyna---- 786  
 QY 1794 NYLTNNPAEALDRINEGIRFEHVNDGQEPVVGGRNGIDSS---ASGKHSVAIGFOKA 1850  
 DB 767 -----gvnanevfkksq-nglhwsgktvngireit--felak 821  
 QY 1851 DGEAAVAIGROTOAGNQSIAIGDNAQATGDSIAIGTGNVAVGKHSNAGIDPSTV--KAD 1908  
 DB 822 dena-----laigyskalrhtlvaigtgnvnaeksgafgdpnyledkag 867  
 QY 1909 NSYSVGNNOFTDATQTDVFGVGN-----ITV-----T 1937  
 DB 868 gsyatgdnrlt--sktflvlgvynakykangdvdtelvtvkdkgkettvtpkalgat 926  
 QY 1938 ESNYSALGSNSAISAGTACGTAQAKKSDGTACTTGTAGTGVKKGAGOTAVGASVGSAG 1997  
 DB 927 vensvylgnk--statckkgknlksdgaqntlaagtlgtvngtaagatavgsavgsag 983  
 QY 1998 AERRIQNVAAGEVASTIDAVNGSOLYKATOGIANATMELDHRHONENKANAGISSAMA 2057  
 DB 984 eerriqnvaaageisatstsdalngsl'yavakgvt-----lagynkvkgkradaglaala 1039  
 QY 2058 MASMPQAVYIPGRSMVTGTGIATHNGGAAVAVGLSKISDNGVWFKINGSADTQGHVGAAG 2117  
 DB 1040 aeqlpqaempgkmswslagsysyggnglaigvstrlsdngkvlrlsgltngsgktygaag 1099  
 QY 2118 AGFHF 2122  
 DB 1100 vgyqw 1104

RESULT 12  
 AAB23859  
 ID AAB23859 standard; Protein: 1104 AA.  
 AC AAB23859;  
 XX  
 DT 17-JAN-2001 (first entry)  
 XX  
 DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K22.  
 XX  
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;  
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
 KW diagnosis; immunogenic; antigen.  
 KW  
 OS Haemophilus influenzae.  
 XX  
 PN WO20005191-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 16-MAR-2000; 2000WO-CA00289.  
 XX  
 PR 16-MAR-1999; 99US-0268347.  
 XX  
 PA (CONN-) CONNACHT LAB LTD.  
 XX  
 PI Logsmore SM, Yang Y, Klein MH;  
 XX  
 DR WPI: 2000-618897/59.

DR N-PSDB: AAA92498.  
 XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
 PT use as antigens and vaccines and for treating Hemophilus influenzae  
 PT infection  
 PS Claim 1; Fig 23; 275pp; English.  
 XX  
 CC The present sequence represents a Haemophilus influenzae adhesin (Hia)  
 CC protein from the non-typeable Haemophilus influenzae (NTH) strain K22.  
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
 CC activities, and can be used in the production of a vaccine. An  
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
 CC protection against disease caused by Haemophilus strains in a  
 CC susceptible host, preferably a human. An Hia protein is useful as an  
 CC antigen, in immunogenic preparations including vaccines, as a carrier  
 CC for other immunogens, and in the generation of diagnostic reagents. Hia  
 CC is useful for treating diseases caused by the infection of Haemophilus  
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis  
 CC media. Recombinant production of Hia favours high recovery of the  
 CC protein compared to the low recovery of native protein from Haemophilus  
 CC influenzae species. A truncated protein has a significantly higher  
 CC amount of recovery than a full-length protein.  
 CC  
 XX Sequence 1104 AA:  
 SQ  
 Query Match 9.5%; Score 1018.5; DB 21; Length 1104;  
 Best Local Similarity 27.8%; Pred. No. 9.1e-36;  
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;  
 QY 943 KASVAVDVNDGTTIHLTGADGNKNOIGVTTTLTKTDKADKAINFSVNSGDD----- 997  
 DB 28 kcaatavavlatatlatat-aeann-----tsvtnglmgysdntfntmstadielkv 81  
 QY 998 ----KALINAKD-----IADNLTLAGLR-----NKGADTALQTFQYKVK 1037  
 DB 82 qdaykgljlnlekdtnksfivadtaevnrlkglvslskngtrne--ksyqvqad 139  
 QY 1038 ENGDDDDADDTIYVGNKAKTQVNTLKLKGNGLDIOIKNNDGYTFPFINOSGKAGNNT 1097  
 DB 140 e-----vltfsgsaatvs-----ssskdkgkhtllsvtkgsfaevkt 176  
 QY 1098 TLNNGSLIKNTAGNEIOVAGDGVKFAKYNNGVAGIGITRTRTFDEIGFAGTNGSLD 1157  
 DB 177 datgq-----qynadgkykaeden--gadvd----- 202  
 QY 1158 KSKPHLSKDGINAGKKTITNIOGSEIAONSNDVATGKRIYDLKTELENKISSAKTAQNS 1217  
 DB 203 -----kkvatvk--dvakaindaatfvtvestdddlengagaqnetldga 245  
 QY 1218 LHEFSVADEGNNFTY---SNPYSSYDTSKTSVITTFAGENGITTKNKGAVRGIDQTK 1274  
 DB 246 l-----kagddcltlkagknlkxldqngksvtfalakldvtsakvsdklsigkdnk 298  
 QY 1275 -----GTTTPTLVGNNGKGIIVDSONGNTITGLSNTLANVTNDSGVRTEQOK 1326  
 DB 299 vdltsdanglkakl--tgnng-----qngvnhlngiasltl--tdltgmmtqasn 345  
 QY 1327 IIKEDKTRAASTIVDSAGFNLOGEAVDESVDTVTVPADGNMTATKAVTYDDTSKTS 1386  
 DB 346 gvaqgmhnaasavdlnagwmldqngasvdfvnaydltvfngtnlvuttdclahkt 405  
 QY 1387 KVVVD-----VNVDDTFTEVK-----DKKIG-----VKTTLTSTGNG 1419  
 DB 406 tvrvdvrglpyqvvtedgkvtvkvdkyyeakqgsadmkkvengelaktkxklysaag 465  
 QY 1420 ANKFALSNQATGALVKSADIVA--HLNLTSGDIQTAKGASQANSAGYDADGNKYI-- 1475  
 DB 466 qnpvklsnvaeg---teendavsfkqlkalq--ekyvtlltasnayangn--daagykacqt 520  
 QY 1476 -----YSTDNKRYQAK--ND-----GTVDKTKEVAKDKLVAAQNPDPGTLAOK 1517

DB 521 lnglnfkfstldgejllnkqvendvtfcpkgsv--qvgedgkatqngtktdtdgveas 579  
 QY 1518 NVKASVINKEDVNDANKKOGINEDNAFVKGLERKASDNKTKMAAVYGDLANAAQPLPFA 1577  
 DB 580 el-----veslnklgwkvyvdkgds-----geldgasnetlvs 613  
 QY 1578 GDTGTTAKKLGELTLTKGGOTDPTNKLTDNNIGVACTDGEFVYKAKDLNLSV----- 1631  
 DB 614 gdvvll--kagelnkvk--qdytn-----ftyalldelgvxsvelfkda 654  
 QY 1632 ---NAGGTRIDKGVSPVDSG-----QAKANTPYLSANGLDVGKVTISNKGKTRKDA 1683  
 DB 655 nngsgastktdqyltltasangangaatdcklkvasdgsiaqgnkavnnvsgjklkfyd 714  
 QY 1684 ANYQULEVNRNL-----GLGNAGDNADGNVNIADIKKPDNSGSSSRVYKAG 1734  
 DB 715 anfnltsadnltkyddaykyltnldekgackqfllvad-----nt-----aa 759  
 QY 1735 TVVGGKGNNDTEKLVAGVGVVDKDGKGNANGLDSNV--WVKTKDKSKKALLATYNAAGT 1793  
 DB 760 lv-----gdltglgw--lsadktlgeInkeyna---- 786  
 QY 1794 NYLTNPAPALDRINQGRFPHVNDGNOEPVVOGRNGIDSS---ASGKHSVAIGFOAKA 1850  
 DB 787 -----qvranevkkfksq--nglhvsgktvngtrfellt--felak 821  
 QY 1851 DGRPAVAIGRQVQAGNOSTAIGDNQATGDSIAIGTGNVAVKAGISAGIPSTV--KAD 1908  
 DB 822 dena-----lafygskalrndeivaigtvnnvnaeksgafgdpnyledkag 867  
 QY 1909 NSYSVGNNOFTDQTQDVEFGVGN-----ITV-----T 1937  
 DB 868 gsyafgndnrlt--skrtflngvnyakykangdvdeitvtydkdkgkeltvvpkajal 926  
 QY 1938 ESNVSALGSNSAISACTHAGTQAKRSDDTAGTTTATAGATGVKGFAGOTAVAGSVASG 1997  
 DB 927 vensvlygnk-----statkdkyknlkedgtagntttagttgvtngfagstlabavagsag 983  
 QY 1998 AERRIONVAAGEVSATSTAVANGSOLYKATOGIAANTNELDRIRIONEKKANAGISSAMA 2057  
 DB 984 eerrignvaageisatstaidangsqlyavakgvtn-----laqvnykvkkrirdagfasala 1039  
 QY 2058 MASMPQAVIPGRSMVTGATATNCGOAVAVGSKLSDNGOWFKINGSDADQGHGAIVG 2117  
 DB 1040 asqldpqaampqsmwsiaagsyqgqnglaivysrdsngkvliritsgtmsgkvtcvaag 1099  
 QY 2118 AGFHF 2122  
 DB 1100 vgyqw 1104  
 RESULT 13  
 AAB23857  
 ID AAB23857 standard; Protein; 1004 AA.  
 XX  
 XX AAB23857;  
 AC  
 XX  
 XX  
 DT 17-JAN-2001 (first entry)  
 XX  
 XX Haemophilus influenzae adhesin (Hia) protein from NTH strain M407.  
 DE  
 XX Hia; adhesin; Haemophilus influenzae adhesin; NTH; infection; vaccine;  
 KW non-typeable; Haemophilus influenzae; antiinflammatory; auditory;  
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;  
 KW diagnosis; immunogenic; antigen.  
 XX  
 XX Haemophilus influenzae.  
 OS  
 XX  
 XX WO200055191-A2.  
 XX  
 XX 21-SEP-2000.  
 PD  
 XX







PR 19-SEP-1996; 96JP-0271408.  
XX (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.  
PA (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.  
XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;  
XX MPI, 1998-230318/20.  
DR N-PSDB; AAV22837.  
XX  
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI  
XX antibody production - and is useful for diagnosis of and preparation  
XX of vaccines for chicken infectious coryza  
PS Claim 5; Page 71-87; 108pp; Japanese.  
XX  
XX The present sequence represents an antigenic protein derived from  
CC Haemophilus paragallinarum strain C-53-47. The antigenic protein  
CC stimulates the production of HI antibodies in fowl. The protein  
CC and DNA coding for it can be used in the preparation of vaccines  
CC for the prevention of chicken infectious coryza (CIC). The protein  
CC and its antibodies can be used in the diagnosis and treatment of CIC.  
XX  
SQ Sequence 2039 AA;  
  
Query Match 7.6%; Score 809.5; DB 19; Length 2039;  
Best Local Similarity 21.6%; Pred. NO. 1.6e-26;  
Matches 515; Conservative 301; Mismatches 830; Indels 735; Gaps 102;  
  
OY 12 ATGTMAVAEYKSHSTGGSCATGCVSVRLSPARIALAVLYIGATLNGSAVAGIGI 71  
DB 124 atgtmavaeykshstggscatgcvsvrlsparialavlyigatlngsavagigi 177  
OY 72 ---SEADGGKGGANARQDKSIAIGDIAQALGQSIAIDNKIVHNSNNANIGAKASGNE 128  
DB 178 netytqldprapessrg---svvigenakssegvslgpn-----awaktn 221  
OY 129 STAIGSDVLAASHASIAIGSD-----STQVN---LNRGIALGFGSOVLQKDNVNA 155  
DB 222 stisagatfaegkstiaigtckilgtyncklpapswdrtgkaplnsiwdfiselymgk 281  
OY 156 ETVOQISELPIIRGOKALNDIYOLADNLQKY-----RRTTHAG 195  
DB 282 kngcdydakkkndrpnkpeafytsdfk-sryvnpstsplyaaiklaialagstliaag 340  
OY 196 HASTAVGANASYAKGHFNAFTRA-----TAEG---TY---SLAVGLTATAKAAS 239  
DB 341 emstavgslafadckstamjlrstfvakdaygtalgeestrfakdsvalgnktaasnag 400  
OY 240 STAVGSMAOALGFAATAVG-----STQVN---LNRGIALGFGSOVLQKDNVNA 287  
DB 401 smaygkakavagalaigaevaagaefdsqagalllnrfgyaliks--adksddikag 458  
OY 288 N-VRAVAPDNDPINDRRKATPEKNGATDVESIGNSGNDSTIRKLIINAGASADIDVAV 346  
DB 459 dainvftgfdmftqshlyentllyltsagd-----lkktaavgdgg----- 504  
OY 347 AOLKPAVRLANRQITFKGDDSNRRVEKGLGTLTTGGQTSALDHNHNGVONGDLKV 406  
DB 505 ---knaiaignk-----taskanvalasysalaesgn----- 534  
OY 407 QLAETLTSIKVNTTLENLANEKEV- VGRTRLTTRDKIGFTNDMNGIDESKPYLDKDTGHA 465  
DB 535 ---afalgysylvsp--laentliavg-----gyatgsnsf-----v 567  
OY 466 GGOKITKLKLAGVY-----DDDAATYGQLKKVNOJATASALQTTVKKVKNDA 514  
DB 568 ggsawstclartvlygsssisdshdslamgvnafignsnssialgtstliakn--a 624  
OY 515 NSKSIITVGNKKNKPDQTVNLTCLKGENGV-----DVTTETNT-----YTF 556  
DB 625 kpsdslaigkdsridakdt-----dngvlyltpgyvdecttraftfdeknkymqamal 677

OY 557 GLNQN-----NGLTVG---NSTLND---GLSVKNTNSNKOIOVGADITFD 598  
DB 678 gfnakvrygkyketyjnsmaigaarsqatlgnstalgvaktylwegle-----ad 729  
OY 599 ISNSKPGAGIENTRITRIRDIGFANN-----TGSIDANKRPLRPTTGIMAGKELTNV 650  
DB 730 pwekagaisiptgkligvisvsgyserrilvnaasgldtd-----avna--qlkcl 780  
OY 651 ---QSAINPATNGQIDFNKRLSTANTEKSGSAAITKDLINLSQVPLRPAOTGRVAVK 706  
DB 781 eertfseidllnggygyqy---lsvektlngeagrv-----asq----- 818  
OY 707 KLGE-----ILKVGKRTTADRLKNNIGVADSTNLSFLVKLKTTSDIDAVNTKLTAS 762  
DB 819 rkgesykrvylktqlllyldarkln-----gekfdqtsldiskavgeaesygel--- 871  
OY 763 DKVTVDSGNMTAKLONDLT-----FSKONTGATPATNSKTIGVDL-----KFT 807  
DB 872 -ktaselnrvanglnaetvndtfgkfykftjenaatnadsekng--glspqvlaqlk 928  
OY 808 DNGIALDGTYYTKDKVGRPAKODGSLDKSPYLDKDKLVG-----EVEITTING- 857  
DB 929 annnylndgak--gqdsiafvgaktseannglagkqalaifqansaaenalsiglnsd 986  
OY 858 -----INAGKALITGLSNTLIDATNATF-----GHWYOLG 887  
DB 987 tsmtgavalaigkatvtagpksialgqdsstvanalsitssvminglftmfsgspely 1046  
OY 888 FVDSIDTRRAASIDVYNAGNLKNNGDAKDFVSTYTPVFINGNATTAATVYDGRASKV 947  
DB 1047 vlsigtgkerkynva-aggdsqtslealngslyltnlmklagsvntngnann-1 1104  
OY 948 AYDVNDVDTTHLTLGADGNKQIGVKTTLTKTDKADKAINEFSVNSGDKALINAKDIA 1007  
DB 1105 atdgtlftnngyggqtdlnhai---nvaltkl-----lslateeevvsgeavy 1152  
OY 1008 DNLN-----TLAGEI-RMTKGADTALOTFOVKYK-----ENGDDNDADITVGRKAKT 1057  
DB 1153 dalgkaprvsaeankgitglvdy-----vkkanspietvpsldmkkkftfvgllakdl 1206  
OY 1058 HQVNTLKLKGNKGDIDOTNKGVTFTGINTQSG-LKAGNNTTLNNGLSIKRTAGNEQ-- 1114  
DB 1207 egvnsitf-dksqgdlnvgvgrmssaglfkfgdttgsttlttaedgltldsttnsqtn 1265  
OY 1115 -IYVGADGVKFAKVNQ-----VVGAGIDGTTTRTRDEIGFAGTNGSLDKSPH 1162  
DB 1266 lvksrdgf---svknsgdsdeklasklsigaenaehvetksjalakadnts-dksait 1321  
OY 1163 LSKRGINAGK-----KITMQGEIQAONSNDNAVTEGKIYDL-----KTEL 1203  
DB 1322 lagaditlagatgalkltgvaagnltvnskdavngqqlrtllgyvsgakigtteftl 1381  
OY 1264 ENKISSTAKTQNSLHEFSVADDEOGNFTVSNPSSVDTSTSPVTFPAGENGITTVK-N 1262  
DB 1382 seaisdvkqaltatlaykaankng-----tkkltdglnftstnldasved 1429  
OY 1263 KGVVRVGI-DQTKGLTTPPKLTVGNNGNGGIYDSQN-GONTITGLSNTLANVTNDKSVR 1320  
DB 1430 ngvvkftklkdlgtlkt-----lateslnsqnliagvtvtygget-egivl 1475  
OY 1321 TTEOGKIITKEDKTRAAISYDVLSAGFNLQNGCAVDFVSTYDIPVNRADGNATTAKYTD 1380  
DB 1476 t-----ksgsngdrfls-----lsagnaatdglkvsavka-----gladdavanng 1517  
OY 1381 DTSKTSKVYVVDVDDTIEVKDKKLGKVTTLTSTGFGANKFALSQATGADALVKASDI 1440  
DB 1518 gldklfkaina-----lg--ttldavt-knpngesirfnpingtepttkda 1561  
OY 1441 VAHLNLTSGDIQTAKGASQANSSAGY--VDADGNKRVITYDSTDNKRYQAKNDGIVDKTREV 1498  
DB 1562 vdklilta---vntgwsavglatgldgda-gukkslnvadgdlspstsgd----- 1608



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:52 ; Search time 33.83 Seconds  
(without alignments)  
1532.105 Million cell updates/sec

Title: US-09-813-214A-9  
Perfect score: 10708  
Sequence: 1 MNHLYKVFENKATGTFMAYV.....NCSADTQGHVCAAGAGFHF 2122

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10697.5	99.9	2123	US-08-968-685A-10	Sequence 10, Appl
2	7320	68.4	2048	US-09-268-347-48	Sequence 48, Appl
3	6585	61.5	2314	US-09-268-347-49	Sequence 49, Appl
4	1185.5	11.1	2353	US-09-377-155-33	Sequence 33, Appl
5	1185.5	11.1	2353	US-08-913-942-4	Sequence 4, Appl
6	1185.5	11.1	2353	US-09-669-974-33	Sequence 33, Appl
7	1177	11.0	2354	US-09-268-347-47	Sequence 47, Appl
8	1169.5	10.9	2411	US-09-268-347-36	Sequence 36, Appl
9	1018.5	9.5	1104	US-09-268-347-28	Sequence 28, Appl
10	1018.5	9.5	1104	US-09-268-347-34	Sequence 34, Appl
11	974	9.1	1004	US-09-268-347-30	Sequence 30, Appl
12	944.5	8.8	1002	US-09-268-347-24	Sequence 24, Appl
13	845	7.9	1912	US-08-409-995-4	Sequence 4, Appl
14	845	7.9	1912	US-08-685-467-4	Sequence 4, Appl
15	693	6.5	1094	US-09-268-347-32	Sequence 32, Appl
16	681	6.4	1098	US-08-409-995-2	Sequence 2, Appl
17	681	6.4	1098	US-08-685-467-2	Sequence 2, Appl
18	681	6.4	1098	US-09-377-155-32	Sequence 32, Appl
19	681	6.4	1098	US-08-913-942-2	Sequence 2, Appl
20	681	6.4	1098	US-09-669-974-32	Sequence 32, Appl
21	681	6.4	1098	US-09-268-347-44	Sequence 44, Appl
22	582.5	5.4	679	US-08-913-942-15	Sequence 15, Appl
23	582.5	5.4	679	US-09-268-347-26	Sequence 26, Appl
24	542	5.1	1600	US-08-617-697-10	Sequence 10, Appl
25	525.5	4.9	1599	US-08-617-697-9	Sequence 9, Appl
26	509	4.8	1536	US-08-038-682-2	Sequence 2, Appl
27	509	4.8	1536	US-08-302-832-2	Sequence 2, Appl

28	509	4.8	1536	2	US-08-530-198-2	Sequence 2, Appl
29	509	4.8	1536	2	US-08-469-880-2	Sequence 2, Appl
30	509	4.8	1536	2	US-08-728-470-2	Sequence 2, Appl
31	509	4.8	1536	2	US-08-617-697-2	Sequence 2, Appl
32	509	4.8	1536	4	US-08-719-641-2	Sequence 10, Appl
33	509	4.7	1529	4	US-08-728-470-10	Sequence 10, Appl
34	500	4.7	1529	4	US-08-719-641-10	Sequence 10, Appl
35	487.5	4.6	1477	1	US-08-038-682-4	Sequence 4, Appl
36	487.5	4.6	1477	1	US-08-302-832-4	Sequence 4, Appl
37	487.5	4.6	1477	2	US-08-530-198-4	Sequence 4, Appl
38	487.5	4.6	1477	2	US-08-469-880-4	Sequence 4, Appl
39	487.5	4.6	1477	2	US-08-728-470-4	Sequence 4, Appl
40	487.5	4.6	1477	2	US-08-617-697-4	Sequence 4, Appl
41	487.5	4.6	1477	4	US-08-719-641-4	Sequence 4, Appl
42	461	4.3	1338	2	US-08-728-470-9	Sequence 9, Appl
43	461	4.3	1338	4	US-08-719-641-9	Sequence 9, Appl
44	458.5	4.3	1612	1	US-08-169-927-2	Sequence 2, Appl
45	401.5	3.7	599	4	US-09-377-155-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1  
US-08-968-685A-10  
Sequence 10, Application US/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:  
APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSLIA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2123 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-10  
Query Match 99.9% ; Score 10697.5 ; DB 4 ; Length 2123 ;  
Best Local Similarity 100.0% ; Pred. No. 0 ;  
Matches 2122 ; Conservative 0 ; Mismatches 0 ; Indels 1 ; Gaps 1 ;  
1 MNHLYKVFENKATGTFMAYVAKSHSTGGSCATGCGVRLTSPFARIALAVLVIGAT 60  
|||||

Db 1 MNHIIYKFNKATGTFMAVAEYAKSHSTGGSCATQOVGSVRTLSEFARIAALAVIGAT 60  
 QY 61 LMSAAVAGIGISEADGKGANARGPDKSIAIGDIAOLGSSOSTAIGDNKIVHSSNNAN 120  
 Db 61 LMSAAVAGIGISEADGKGANARGPDKSIAIGDIAOLGSSOSTAIGDNKIVHSSNNAN 120  
 QY 121 GAKASGESIAIGDVLASGASIAIGSDLYLKEETVOOISELPTIRGOKALNDIYL 180  
 Db 121 GAKASGESIAIGDVLASGASIAIGSDLYLKEETVOOISELPTIRGOKALNDIYL 180  
 QY 181 ADNLQKRRTHQOHAHSTAVGAMSTAKGHFSNAPGTBRATBEYSILAVGLTATAAAS 240  
 Db 181 ADNLQKRRTHQOHAHSTAVGAMSTAKGHFSNAPGTBRATBEYSILAVGLTATAAAS 240  
 QY 241 IAVGNAOAGIAPAVGSGTQVNLNRGIALGFSQVLOKNDVNAANRATAPDNOPI 300  
 Db 241 IAVGNAOAGIAPAVGSGTQVNLNRGIALGFSQVLOKNDVNAANRATAPDNOPI 300  
 QY 301 DNRKATFKNGATVFSIGNSNGDSIRKIIIVGAGSADTDVAVNAOLKEAVRLANRO 360  
 Db 301 DNRKATFKNGATVFSIGNSNGDSIRKIIIVGAGSADTDVAVNAOLKEAVRLANRO 360  
 QY 361 TFGDSSNNNEVEKGLKTLITGGAQTSALTBNIGVONGDLKQOLAEITLSLMTT 420  
 Db 361 TFGDSSNNNEVEKGLKTLITGGAQTSALTBNIGVONGDLKQOLAEITLSLMTT 420  
 QY 421 ENLTANEKYVCKRLTDTKIGFTNDMNGIDESKPYLDKDTGIAHAGOKITKLTAGVND 480  
 Db 421 ENLTANEKYVCKRLTDTKIGFTNDMNGIDESKPYLDKDTGIAHAGOKITKLTAGVND 480  
 QY 481 DAATYQOLKVNQATASALQTFVKKVKNKNDANDANSKIIIVGKNNKPDGTQVNTLKLKG 540  
 Db 481 DAATYQOLKVNQATASALQTFVKKVKNKNDANDANSKIIIVGKNNKPDGTQVNTLKLKG 540  
 QY 541 ENGVDYTFENGTVEGMLNONNGLTGNSSTLNDGSLVKNNTSNNOIQVGAAGITPTDIS 600  
 Db 541 ENGVDYTFENGTVEGMLNONNGLTGNSSTLNDGSLVKNNTSNNOIQVGAAGITPTDIS 600  
 QY 601 NSKPGAGIENTTIRTRDQIGFANNTGSLDANKPRLPTGINSAGKELTIVGSAINPATNG 660  
 Db 601 NSKPGAGIENTTIRTRDQIGFANNTGSLDANKPRLPTGINSAGKELTIVGSAINPATNG 660  
 QY 661 GOLDFENRISTANTEKSGSAAITKDLYNLSQVPLTFAGDTGPNVTKLGEIILKVGKTT 720  
 Db 661 GOLDFENRISTANTEKSGSAAITKDLYNLSQVPLTFAGDTGPNVTKLGEIILKVGKTT 720  
 QY 721 ADDLTNNNGVAVADSTNSLTIVKLAITSDLDAVNTKTLTASDKVTVDSGNTFAKLQND 780  
 Db 721 ADDLTNNNGVAVADSTNSLTIVKLAITSDLDAVNTKTLTASDKVTVDSGNTFAKLQND 780  
 QY 781 LFFSKONTGATPATNSKTI-GVDGLKFTDNGIALDGTYYIRKDVYGAFAKOGSILDKSP 839  
 Db 781 LFFSKONTGATPATNSKTI-GVDGLKFTDNGIALDGTYYIRKDVYGAFAKOGSILDKSP 839  
 QY 840 YLDKDKLAKGEVEITTINGINAGKALITGLSNTLDTATNAITGHVTLQGLIVSDTKTRAAS 899  
 Db 840 YLDKDKLAKGEVEITTINGINAGKALITGLSNTLDTATNAITGHVTLQGLIVSDTKTRAAS 899  
 QY 900 IGDVLNAGFNLNKNGDAKFEVSTYDTPVDFINGNATTAKTYGKASKAYAVDVNNGTITH 959  
 Db 900 IGDVLNAGFNLNKNGDAKFEVSTYDTPVDFINGNATTAKTYGKASKAYAVDVNNGTITH 959  
 QY 960 LFGADGNKNOIGVKTTLTKTDKAGKALINFSVNSGDDKALINADIDNLTLTLAGETRN 1019  
 Db 960 LFGADGNKNOIGVKTTLTKTDKAGKALINFSVNSGDDKALINADIDNLTLTLAGETRN 1019  
 QY 1020 TFGTADTALQTFQVKKVENGGDDNDADTIVGKDAKTNOVNTLKLKKGKNGIDITQNKDG 1079  
 Db 1020 TFGTADTALQTFQVKKVENGGDDNDADTIVGKDAKTNOVNTLKLKKGKNGIDITQNKDG 1079  
 QY 1080 TWTFGINTOSGLKAGNNTLNNNGLSIKRTAGNEOIOVGADGVKPAKVNNGVAGAGIDGT 1139  
 Db 1080 TWTFGINTOSGLKAGNNTLNNNGLSIKRTAGNEOIOVGADGVKPAKVNNGVAGAGIDGT 1139  
 QY 1081 TWTFGINTOSGLKAGNNTLNNNGLSIKRTAGNEOIOVGADGVKPAKVNNGVAGAGIDGT 1140  
 Db 1081 TWTFGINTOSGLKAGNNTLNNNGLSIKRTAGNEOIOVGADGVKPAKVNNGVAGAGIDGT 1140

QY 1140 TRTRDEIGFAGTNGSLDKSPHLKSDGINAGKKTITNIOGSEITAOONSNDAVTGKLYDL 1199  
 Db 1141 TRTRDEIGFAGTNGSLDKSPHLKSDGINAGKKTITNIOGSEITAOONSNDAVTGKLYDL 1200  
 QY 1200 KTELENKISSTAKTAQNSLHEFSVADGONNETVSNPSSYDTSKTSVITFAGEENGITT 1259  
 Db 1201 KTELENKISSTAKTAQNSLHEFSVADGONNETVSNPSSYDTSKTSVITFAGEENGITT 1260  
 QY 1260 KYNKGVAVGIDQTKGLTTPKLYGNNGNGKGIIVDSONGONTITGLSNTLANVTNDKGSV 1319  
 Db 1261 KYNKGVAVGIDQTKGLTTPKLYGNNGNGKGIIVDSONGONTITGLSNTLANVTNDKGSV 1320  
 QY 1320 RTTEBGKIIKEDBKTRASIVDVSAGFNLOGNEAVDFESTYDTPVADGNATTATVY 1379  
 Db 1321 RTTEBGKIIKEDBKTRASIVDVSAGFNLOGNEAVDFESTYDTPVADGNATTATVY 1380  
 QY 1380 DDTSKTSKVYDVAVNDPTTIEVKDKKLGKTTTLTSTGTGANKFALSNOATGALVRSAD 1439  
 Db 1381 DDTSKTSKVYDVAVNDPTTIEVKDKKLGKTTTLTSTGTGANKFALSNOATGALVRSAD 1440  
 QY 1440 IVAHLNLTSGDIQTRAKGASQANSAGYVDADGNKYIYDSTDNKTYOAKNDGTVDKTEVA 1499  
 Db 1441 IVAHLNLTSGDIQTRAKGASQANSAGYVDADGNKYIYDSTDNKTYOAKNDGTVDKTEVA 1500  
 QY 1500 KDKLVAQAOTPDGTLAQNANKSVINKBEQVNDANKKOGINEANAVKGLERASDNKTKNA 1559  
 Db 1501 KDKLVAQAOTPDGTLAQNANKSVINKBEQVNDANKKOGINEANAVKGLERASDNKTKNA 1560  
 QY 1560 AVTVGDLNAVAQPTLTFAGDTGTAKKLGELTLTKGGQDTNKLTDNNIGVAVGTDEFTV 1619  
 Db 1561 AVTVGDLNAVAQPTLTFAGDTGTAKKLGELTLTKGGQDTNKLTDNNIGVAVGTDEFTV 1620  
 QY 1620 KLADLTMLNSVNAAGTITDKKGVSPVDSGQAAKANTPVISANGLDGGKVISVVGKGT 1679  
 Db 1621 KLADLTMLNSVNAAGTITDKKGVSPVDSGQAAKANTPVISANGLDGGKVISVVGKGT 1680  
 QY 1680 DTDAAVQOLNEVRNLGLGAGNDNADGNQVNTADIKKPDNSGSSSNRYIRAGTYLGG 1739  
 Db 1681 DTDAAVQOLNEVRNLGLGAGNDNADGNQVNTADIKKPDNSGSSSNRYIRAGTYLGG 1740  
 QY 1740 KGNNDTEKLTAGVQVGVYDKDGNANGDLSNMYVKTQKDSKALLATYNAAGQNTYLTNN 1799  
 Db 1741 KGNNDTEKLTAGVQVGVYDKDGNANGDLSNMYVKTQKDSKALLATYNAAGQNTYLTNN 1800  
 QY 1800 PAEALDRINEGIRFEFHVNDGNQEPVVOGRNGIDSSASGRHSAIGFOAKADEAAVAIG 1859  
 Db 1801 PAEALDRINEGIRFEFHVNDGNQEPVVOGRNGIDSSASGRHSAIGFOAKADEAAVAIG 1860  
 QY 1860 ROTOGANOSIAGDMAATGPOSIAITGNVYVAKHSGAIGDPSTYKADNSYSVGNNOF 1919  
 Db 1861 ROTOGANOSIAGDMAATGPOSIAITGNVYVAKHSGAIGDPSTYKADNSYSVGNNOF 1920  
 QY 1920 TDAOTQDFVGVGNNTIYTESNSVALGSNSAISACTHAGTOAKKSDGAGTTTATAGATVY 1979  
 Db 1921 TDAOTQDFVGVGNNTIYTESNSVALGSNSAISACTHAGTOAKKSDGAGTTTATAGATVY 1980  
 QY 1980 KGFAGQTRAVGASVSGASGAERIRIONVAAGEVSASTDAVNGSOLKRYATOGIAATNELDH 2039  
 Db 1981 KGFAGQTRAVGASVSGASGAERIRIONVAAGEVSASTDAVNGSOLKRYATOGIAATNELDH 2040  
 QY 2040 RIHONENKANAGISSAAMAMSPQATIPGRSMYTGATINNGGCAVAVGSLKSLSDNGOWY 2099  
 Db 2041 RIHONENKANAGISSAAMAMSPQATIPGRSMYTGATINNGGCAVAVGSLKSLSDNGOWY 2100  
 QY 2100 FKINGSADTOGHVGAAGAGFHF 2122  
 Db 2101 FKINGSADTOGHVGAAGAGFHF 2123

RESULT 2  
 US-09-268-347-48  
 ; Sequence 48, Application US/09268347

Patent No. 6335182  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
FILE REFERENCE: 1038-860  
CURRENT APPLICATION NUMBER: US/09/268,347  
CURRENT FILING DATE: 1999-03-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 48  
LENGTH: 2048  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-268-347-48

Query Match 68.4%; Score 7320; DB 4; Length 2048;  
Best Local Similarity 68.1%; Pred. No. 0;  
Matches 1552; Conservative 117; Mismatches 221; Indels 390; Gaps 24;

QY 1 MNHYKVFNKATGTFMAVAEYAKSHSTGGSCATGQYGVSRITSEFARIALAVLVIGAT 60  
DB 1 MNHYKVFNKATGTFMAVAEYAKSHSTGGSCATGQYGVSRITSEFARIALAVLVIGAT 60  
QY 61 LNSGSAVA-----GIGISEADGKGKGNARSGKSTAGIDIAQALQSOSIAGIDNKIVH 112  
DB 61 LNSGSAVAKKTOTKHAIGEONPRRSSTAKADGDBAIAIGENANAQGOAIAISSNKTIV 120  
QY 113 NSNNANNAKAGAGNESIAGIDVYLASGHASIAIGSDLYL-----KKEETVOOISELL 165  
DB 113 NSNNANNAKAGAGNESIAGIDVYLASGHASIAIGSDLYL-----KKEETVOOISELL 165  
QY 121 NSSLDKICTDATGQESIAIGDVYASGDASIAIGSDLLHLDHGNKHPKGLINDL- 179  
DB 121 NSSLDKICTDATGQESIAIGDVYASGDASIAIGSDLLHLDHGNKHPKGLINDL- 179  
QY 166 PIRIQAKLNDIYQALADINLOKRYTHAOGHASTAVAGMSIAKHFSAFETRAETBYT 225  
DB 166 PIRIQAKLNDIYQALADINLOKRYTHAOGHASTAVAGMSIAKHFSAFETRAETBYT 225  
QY 180 --INGHAVALKEIRSSKMDV- KYRRTTASGHASTAVAGMSIAQGFHSAFETRAETASAY 236  
DB 180 --INGHAVALKEIRSSKMDV- KYRRTTASGHASTAVAGMSIAQGFHSAFETRAETASAY 236  
QY 226 SLAVGLTPTAKAASIAVGSNAQAIQFATAVAGSTOVNLKNGIALGFSQVLOKDNVDN 285  
DB 226 SLAVGLTPTAKAASIAVGSNAQAIQFATAVAGSTOVNLKNGIALGFSQVLOKDNVDN 285  
QY 237 SLAVGLTPTAKAASIAVGSNAQAIQFATAVAGSTOVNLKNGIALGFSQVLOKDNVDN 293  
DB 237 SLAVGLTPTAKAASIAVGSNAQAIQFATAVAGSTOVNLKNGIALGFSQVLOKDNVDN 293  
QY 286 AANVR-AAAAPDNDP IDNRKATEKFNKATDVESIGNSNGNSIRKIIINVGAGSADTDAY 344  
DB 286 AANVR-AAAAPDNDP IDNRKATEKFNKATDVESIGNSNGNSIRKIIINVGAGSADTDAY 344  
QY 294 --NSRPATTP-NTQALDPKFOATNNTKAGPL-----SIGNSIKRKIIINVGAGVAKTDAY 345  
DB 294 --NSRPATTP-NTQALDPKFOATNNTKAGPL-----SIGNSIKRKIIINVGAGVAKTDAY 345  
QY 345 NVAOLKEAVRLA-NQOTFEKGDSDNNREKGLGKLTITGGAQISALTDHITGVQOND- 402  
DB 345 NVAOLKEAVRLA-NQOTFEKGDSDNNREKGLGKLTITGGAQISALTDHITGVQOND- 402  
QY 346 NVAOLEAVYKAKERRITFOGDDNSTDVKIGLDNLTITKGAEETNALDNNIGVYKEDN 405  
DB 346 NVAOLEAVYKAKERRITFOGDDNSTDVKIGLDNLTITKGAEETNALDNNIGVYKEDN 405  
QY 403 -GLKQALLETITSLKMTTEMLTANEKVTGK----- 433  
DB 403 -GLKQALLETITSLKMTTEMLTANEKVTGK----- 433  
QY 406 SGLKYLAKTLNLTLEVTITLNTATTIVKVGSSSSTTAELLSDSLFTQPTNGSOSTSKT 465  
DB 406 SGLKYLAKTLNLTLEVTITLNTATTIVKVGSSSSTTAELLSDSLFTQPTNGSOSTSKT 465  
QY 434 -----TRLTTRKIGFTNDNGIDSKPYLDR-----D 460  
DB 434 -----TRLTTRKIGFTNDNGIDSKPYLDR-----D 460  
QY 466 VYGVNGVFTNNAETTAIGTRITRIRDKIGFARDDVDEKAPAYDKQOLKVGSAITRID 525  
DB 466 VYGVNGVFTNNAETTAIGTRITRIRDKIGFARDDVDEKAPAYDKQOLKVGSAITRID 525  
QY 461 TGIHAGGQKITKLTAGVDDDAATYGQL----- 489  
DB 461 TGIHAGGQKITKLTAGVDDDAATYGQL----- 489  
QY 526 NCIDAGNKKISMLAKSSANDAVITLOLKAAPTLNAGAGISVPTETISVDAKSGNVTAP 585  
DB 526 NCIDAGNKKISMLAKSSANDAVITLOLKAAPTLNAGAGISVPTETISVDAKSGNVTAP 585  
QY 490 -----KVNQTESALQPTTVKV 507  
DB 490 -----KVNQTESALQPTTVKV 507  
QY 586 TYNIGVKTTELSNDSGTSPKFSVSGSGTNSLSVTAHLASLYLEKVRKTDALSQSTTVAE- 644  
DB 586 TYNIGVKTTELSNDSGTSPKFSVSGSGTNSLSVTAHLASLYLEKVRKTDALSQSTTVAE- 644  
QY 508 DKNGDANDSKLITVGNKNNKPDGTQVNTLKLKGENGVDTJETNGTVEFGLNQNGLTVG 567  
DB 508 DKNGDANDSKLITVGNKNNKPDGTQVNTLKLKGENGVDTJETNGTVEFGLNQNGLTVG 567  
QY 645 -----EDDDANAIYAKDPTKNAGAVSILKLGKNGLTATRKDDQVTEFGLSDQGLTIG 700  
DB 645 -----EDDDANAIYAKDPTKNAGAVSILKLGKNGLTATRKDDQVTEFGLSDQGLTIG 700  
QY 568 NSTLNDGILSVANTNSKNOIQVADGCTPTDISNKPAGAGIENTRITRIDGIGFANNTGS 627  
DB 568 NSTLNDGILSVANTNSKNOIQVADGCTPTDISNKPAGAGIENTRITRIDGIGFANNTGS 627  
QY 701 KSTLNDGILTVKDT--NRDIOVGANGIKFTNYVNSNPGIGIANTARITR----- 747  
DB 701 KSTLNDGILTVKDT--NRDIOVGANGIKFTNYVNSNPGIGIANTARITR----- 747  
QY 628 LDANKPRLPTGINSKGLTINVSAINPATNGQLDFNRLSTANTERKSGSATIKDLY 687  
DB 628 LDANKPRLPTGINSKGLTINVSAINPATNGQLDFNRLSTANTERKSGSATIKDLY 687

DB 748 ----- 747  
QY 688 NLSQVPLTPAGDTGPVNTKGLLEIKVKGKITTADLTNNIGVADSDNSLTVKLANT 747  
DB 748 ----- 747  
QY 748 LSLDLAVNFKTLASDKVTVDSGNNTAKLQNDLFSKONTGATPATNSKITIGVGLKFT 807  
DB 748 ----- 747  
QY 808 DNNGIALDGTTYITKDKVGFAPKQDSLDKSKPYLDRKLVGEVEITTINGINAGKALIG 867  
DB 748 -----DKIGRSGSDGAVDTRKPYLDDDKLQYGVKLTNNGINAGKALIG 792  
QY 868 LSNLTDAINATGHTVTOIGIVDSTDKTRAASIGDVLANAGFNLNKNGDAKDEYSTDYD 927  
DB 793 LSPTLPSIADQSSRIEELGNTIDDKSMAASINDLNLGFMKNNNNPIDFVSTYDIYD 852  
QY 928 FINGNATTAQVTD--GKASKVAVDVNDGTTIHLGADGNKNOIGVKTTLTKTPAKGD 985  
DB 986 KAINFSVSGDDKALINAKDIADNLNTLAGEIRNTRFGTADTALQTPQVKKVKEGDDND 1045  
QY 912 EATNFVNSSDEDAVNAKDIADNLNTLAGEIRNTRFGTADTALQTPQVKKVKEGDDND 970  
DB 1046 ADPTVGGKDKKTNQVNTLKLKNGKGLDIOINKDGYTFEGINTOSGLKAGNNTLNNNGLS 1105  
QY 971 AMATVQKRNANNQVNTLKLKNGKGLDIOINKDGYTFEGINTOSGLKAGNNTLNNNGLS 1029  
DB 1106 IKNTAGNEOIQVGDGKFAKV--NNGVYAGIDGTRITRITDELFAGTNGSLDKSPHLS 1164  
QY 1030 IKNTAGNEOIQVGDGKFAKV--NNGVYAGIDGTRITRITDELFAGTNGSLDKSPHLS 1089  
DB 1165 KDGINAGKKITNIOSEIQAONSNDAYTGKIDYDKTELENKISSTAKTAQNSLHEFSVA 1224  
QY 1090 KDGINAGKKITNIOSEIQAONSNDAYTGKIDYDKTELENKISSTAKTAQNSLHEFSVA 1149  
DB 1225 DEOGNNTVSNPSSYDTSTSDYITTFAGENGITTYKKNVAVRGIDQTKGLTPPLTVG 1284  
QY 1150 DEOGNNTVSNPSSYDTSTSDYITTFAGENGITTYKKNVAVRGIDQTKGLTPPLTVG 1209  
DB 1285 NNNKGIIVDSONGQNTITGISMNTLAVTNDKGSVTRTEGKTIKEDDKTRAASIDVLS 1344  
QY 1210 NNNKGIIVDSONGQNTITGISMNTLAVTNDKGSVTRTEGKTIKEDDKTRAASIDVLS 1269  
DB 1345 AGFNLOGNGEAVDFSVTYDVFNFADGNATYATYDTSKTSKRVYVDVNDPTTIEVKD 1404  
QY 1270 AGFNLOGNGEAVDFSVTYDVFNFADGNATYATYDTSKTSKRVYVDVNDPTTIEVKD 1328  
DB 1405 KLGKVTTLTSTGTGANKKALSNQATGDALVYKASDIYAH-LNLSGDIQAKASQANS 1463  
QY 1329 KLGKVTTLTSTGTGANKKALSNQATGDALVYKASDIYAH-LNLSGDIQAKASQANS 1388  
DB 1464 AGVYDADGNK-VIYDSDNTKYOKKNDGTVDKREYAKDLVQAQTPDSTLQOMNVKSV 1522  
QY 1389 AGVYDADGNK-VIYDSDNTKYOKKNDGTVDKREYAKDLVQAQTPDSTLQOMNVKSV 1448  
DB 1523 INKEQVNDANKKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNVAQPTLPFAGDTGT 1582  
QY 1449 INKEQVNDANKKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNVAQPTLPFAGDTGT 1508  
DB 1583 TARKLGETTLTKGCTTNNKLTNNICVAVAGTGFYKLAIDLTLNLSVAVAGTKTIDDKG 1642  
QY 1509 TARKLGETTLTKGCTTNNKLTNNICVAVAGTGFYKLAIDLTLNLSVAVAGTKTIDDKG 1568  
DB 1643 VSFVSSGQAKANPVPVANGLOLGKVIISVNGKGTDTDAANVOOLENFRNLGLGNAG 1702  
QY 1669 VSFVSSGQAKANPVPVANGLOLGKVIISVNGKGTDTDAANVOOLENFRNLGLGNAG 1628  
DB 1703 NDNDAGNOVNTADIKKDPNSGSSSNRTVIRAGTVLGKGNNDTEKLATGGOVGVYDKN 1762  
DB 1703 NDNDAGNOVNTADIKKDPNSGSSSNRTVIRAGTVLGKGNNDTEKLATGGOVGVYDKN 1762

Db 1629 NDNADGNQVNIADIKDPNKGSSNRTVIKAGTVLGGKGNNDTEKLATGGIGVGVDDKGN 1688  
Qy 1763 ANGDLNVMVKTQDKSKKLLATYNAAGOTNLTNNPAEADIRINEOGIRPHVNDGNQ 1822  
Db 1689 ANDLSVWVKTQDKSKKLLATYNAAGOTNLTNNPAEADIRINEOGIRPHVNDGNQ 1748  
Qy 1823 EPVVOGNGIDSSASGKSHVAIGFOAKKADGEAAVAIGROTOAGNOSIAGIDNAQATGDS 1882  
Db 1749 EPVVOGNGIDSSASGKSHVAIGFOAKKADGEAAVAIGROTOAGNOSIAGIDNAQATGDS 1808  
Qy 1883 IAGTGNVAGKSHGAIIGDPSTVKADNYSVGNNOFTDATTOTDVEGVGNNTVTESNSV 1942  
Db 1809 IAGTGNVAGKSHGAIIGDPSTVKADNYSVGNNOFTDATTOTDVEGVGNNTVTESNSV 1868  
Qy 1943 ALGNSAISAGTHAGTQAKKSDGTAGTTAGATGTGKAGGOTAVGAASVAGSAGERRI 2002  
Db 1869 ALGNSAISAGTHAGTQAKKSDGTAGTTAGATGTGKAGGOTAVGAASVAGSAGERRI 1928  
Qy 2003 QNVAGEVSATSTDAVNGSOLYKATOGIANATVELDHRITHONENKANAGISSAMASMP 2062  
Db 1929 QNVAGEVSATSTDAVNGSOLYKATOGIANATVELDHRITHONENKANAGISSAMASMP 1988  
Qy 2063 QAVIPGRSMVTGAIATHNGGAVAVLSKLDNGQVVEFKINSADTQGHVGAAGAGFHF 2122  
Db 1989 QAVIPGRSMVTGAIATHNGGAVAVLSKLDNGQVVEFKINSADTQGHVGAAGAGFHF 2048

## RESULT 3

US-09-268-347-49  
; Sequence 49, Application US/09268347  
; Patent No. 6335182  
; GENERAL INFORMATION:  
; Applicant: Loomis, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347  
; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 2314  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-268-347-49

## Query Match

Best Local Similarity 61.5%; Score 6585; DB 4; Length 2314;  
Matches 1460; Conservative 137; Mismatches 280; Indels 682; Gaps 40;

Qy 1 MNHIVYVFNKATGTFMAVAEYAKSHSTGSGCATQGVSVRFLSPFARIALAVIGAT 60  
Db 1 MNHIVYVFNKATGTFMAVAEYAKSHSTGSGCATQGVSVRFLSPFARIALAVIGAT 60  
Qy 61 LNSGAVAGIGISEADGKG--GANAAGDKSIAIGDIAALGOSISIAIDNR-IVHNSNN 116  
Db 61 LNSGAVAGIGISEADGKG--GANAAGDKSIAIGDIAALGOSISIAIDNR-IVHNSNN 116  
Qy 117 NANIGAKAGNESIAGIGVYLASGHAISIAIGSDLLYKREYQ--ISELLPIIRGQKA 173  
Db 121 NOKAGSHAGKESIAIGVYLASGHAISIAIGSDLLYKREYQ--ISELLPIIRGQKA 180  
Qy 174 LNDIYVADTN-LQYRTHAGHASTAVGAMSYAKGHSNAGFRTATVEGYSTAVGLT 232  
Db 181 LR--QIRSGNSQKRRRAEGHASTAVGAMSYAKGHSNAGFRTATVEGYSTAVGLT 237  
Qy 233 ATAKAASSTAVGNAOIGFAATAVGSGTQVNLNRCIALGFGSOYLQKDNVNAANVAV 292  
Db 238 AKAEEKYTTAIGSNQAIYVYALGADTRVDLYGIALGYSQL--NNNNNNNNKAY 294  
Qy 293 ADDNDPIONRRKATPKNATDVFSIGNSGNDSTIRKTIINGASADTDVAVNAOLKEA 352  
Db 295 VPEGNSNIKSKAT--GNG--LFSIGSS--TIKRIIINGAGYEDTDVAVNAOLKAV 346

Qy 333 VRLANQITFEKGDSSNNRVEKLGKTLITGG-AQTSALTD-HNIGV-QNGDLKQOLA 409  
Db 347 ENLARKQITFEKGDSSNNRVEKLGKTLITGG-AQTSALTD-HNIGV-QNGDLKQOLA 406  
Qy 410 ETLTSLKAVTTEENLANKKVTYVVK- 433  
Db 407 KLSGLEIVSTKRLASKEVYVSGSNNAELQSGGLFTPTTNSATKTYVGTGDLKFTD 466  
Qy 434 -----TLTDDKIGFTNDMNGIDESKPYLDKDYGIHAGGKITKLAVVDDAAT 484  
Db 467 NSNTALEDTTTRITKDKIGFSNAGAGVDEKPKYLDK- 502  
Qy 485 YQGLKVNQTESALQITFVKKVKNQNDANDSKTIYGNKNNKPDGQVNTKLKGENGV 544  
Db 503 -----KTK----- 505  
Qy 545 DYTTEGTNGVTGELNQNGLIVGNSLTNNDLSYKNT--NSNKOIYGADGIFPTDISNS 602  
Db 506 -----VGNSTLNNGLIYVNTTIGSSNKQIYGADGIRKADYVNV 544  
Qy 603 KPGAGIENTTTRTRDGIKGFANNTGSLDANKPRL-----TPTGINAGKELTN 649  
Db 545 VSNNAKFEFTTRITEBEIEFADADGKVDKSPYLDKQLOVGVKRTKDSGINAGDQKISN 604  
Qy 650 VQSA-----INPATNGOLDPMNRLSTANTEKSSAATIKDLYNLSQV-- 693  
Db 605 VKDATDITDAVYTKQKOVQODADALQSFIR-----DEKGEFTISNYSNGNTPNT 658  
Qy 694 ---LTFAGDTPPNVTKKLGELIKVGGKTADDELTKNNIGVAASTNSLTVKLAKTLD 750  
Db 659 FETITFAGENGSI-----SNDLAKGVVK-----G 684  
Qy 751 LDAVNTKTLTASDKVYVDSGNNTAKLONGLTFESKONTGATPATNSKTIGVDLKTDDN 810  
Db 685 IDPIN-----GLTPEKL-----TVGSD----- 701  
Qy 811 GIALDGTITRKDKVAFKQDGLSKSPYLDKKLKVEYEITNGINAGKATISLN 870  
Db 702 ---KDKTQV-----LEVASG--NGTKNIRKVS 728  
Qy 871 TLTDATNATGHTVQLG--IYDSTDKTFRASIGVYLNAGFNLKNNGDAKDFVSTYDVF 928  
Db 729 TLPSITNNGVRTIEGNTIISDEKSKAASIGDILMTGFLKNNSSVGVSYNVD 788  
Qy 929 INGAATTAATYD--GKASKVAVVNVNDGTTIHLTGADGNKNOIGVTTTLTKTDAGDK 986  
Db 789 IDGNATTAATYDDETNOTSKVYVNVNDEKTIELTGNGKTKNKIGVKTTLTNNANG-K 847  
Qy 987 AINFSVNSGDDKALINKADIDNLTLAGELRNTKGTADTALQFFQYKVKENGDDONDA 1046  
Db 848 AINFSVNSGDDKALINKADIDNLTLAGELRNTKGTADTALQFFQYKVKENGDDONDA 1046  
Qy 1047 DITVYKDKAKTN--QVNTLKLKNGKGLDITQNKDGTTFGINTQSLKAGNNTTLNNG 1104  
Db 901 EITVYKDKAKTN--QVNTLKLKNGKGLDITQNKDGTTFGINTQSLKAGNNTTLNNG 1104  
Qy 1105 SIKNTAGNEQIYGADGVKPKVNNVYVAGCIDTTRITRDEIGFAGTNSLSDKSKHLS 1164  
Db 961 SIKNTAGNEQIYGADGVKPKVNNVYVAGCIDTTRITRDEIGFAGTNSLSDKSKHLS 1020  
Qy 1165 KD-----GINAGKRTITNIOGSEIJAOSNAPAVGKTYDTELENNKISSTAK 1212  
Db 1021 KDKLKGVEITETNGINAGKRTITNIOGSEIJAOSNAPAVGKTYDTELENNKISSTAK 1080  
Qy 1213 TAONSLHEFVADQGNNTVSNPYSSYDTSKTSVDTTFAGENGITTKVNVGVVAGIDQ 1272  
Db 1061 TAONSLHEFVADQGNNTVSNPYSSYDTSKTSVDTTFAGENGITTKVNVGVVAGIDQ 1140  
Qy 1273 EKGITLPRILTYGNNNGGIYVDSONGONTITGLSNTLANVTNDKGSVRTTEQGIKDED 1332  
Db 1141 EKGITLPRILTYGNNNGGIYVDSONGONTITGLSNTLANVTND--GAGHALSOG--LANDTD 1198

QY 1333 KTRAAISIVDVLASAGNENLOGEAVDFVSTYDVTNFEADGNATKAVTYDDTSKTSKVYDV 1392  
 1199 KTRAAISIDVLAGNENLOGEAVDFVSTYDVTNFEADGNATKAVTYDDTSKTSKVYDV 1258  
 QY 1393 NVDOTTEY-KDKLIGVTTTTLSTIGCAN--KTRALSQAATGALVKASDIIVAHNLTLISG 1449  
 1259 NVDNNTIETVTDKRLGKGVTTTTLSTIGCAN--KTRALSQAATGALVKASDIIVAHNLTLISG 1315  
 QY 1450 DIOTAKGASOANSAGYVDADGNVYDSTDNKYYOAKNDSTVKTATKVAADKLVAAQOT 1509  
 1316 DIOTAKGASOANSAGYVDADGNVYDSTDNKYYOAKNDSTVKTATKVAADKLVAAQOT 1375  
 QY 1510 PDGTLAQNNAVSIVKEQVNDANKKOGINEDNAFVKELEKASDNKTNAATVAGDLNAV 1569  
 1376 PDGTLAQNNAVSIVKEQVNDANKKOGINEDNAFVKELEKASDNKTNAATVAGDLNAV 1455  
 QY 1570 AOTPLTFAGDGTAKKIGETLITKGGOTDNNKLTNNIGVAGTGGTVAKLADLTNLN 1629  
 1436 AOTPLTFAGDGTAKKIGETLITKGGOTDNNKLTNNIGVAGTGGTVAKLADLTNLN 1495  
 QY 1630 SVNAGGTIDDKGVSEFVDSGOAKNPVLSANGLD----- 1666  
 1496 SVNAGGTIDDKGVSEFVDSGOAKNPVLSANGLD----- 1555  
 QY 1667 ----- 1666  
 1556 NEVAKTVNNLNQNSGASLPFVYTDANGKPIINGTDGKPKQAIKAGADKTYHANANGVPV 1615  
 QY 1667 ----- 1666  
 1616 DKDKPITDADKLANIAHGRPLDAGHQVVASLGSNSDAITLTNIKSTLPQIDPTPTGNA 1675  
 QY 1667 ----- 1666  
 1676 NAGQAQSLPSSLSAAQOASNAASVKDVLNAGVFNLOQNHQVDFVAKYDTVNFVNGTADITS 1735  
 QY 1667 ----- 1666  
 1736 VRASDGTMSNTVTMTALATDODGNVLKADGKFKYKADLMPKSLKAGSASADATPT 1795  
 QY 1667 ----- 1666  
 1796 GLSLVNPAGSGTGDVAVALNLSKAVFKSKDGTITTTTSSDGISIOGKNSITLSKG 1855  
 QY 1667 ----- 1666  
 1667 ---GGKVSNNYKGTGKDDANNOOLNEVRNLLGLGNAAGNDNAGNOVNTADIKDPNSG 1723  
 1856 LNVGKVSNNYKGTGKDDANNOOLNEVRNLLGLGNAAGNDNAGNOVNTADIKDPNSG 1915  
 QY 1724 SSSNRTVIAKGTVLGKGNNDTEKLTATGVOYVDKDNAGNDLSNWKTKQKDSKKAL 1783  
 1916 SSSNRTVIAKGTVLGKGNNDTEKLTATGVOYVDKDNAGNDLSNWKTKQKDSKKAL 1975  
 QY 1784 LATYNAAGQTVNLNNPAEALIDRINEGIRPFHVNDGNOEFPVYOGKNGIDSSASGKHSVA 1843  
 1976 LATYNAAGQTVNLNNPAEALIDRINEGIRPFHVNDGNOEFPVYOGKNGIDSSASGKHSVA 2035  
 QY 1844 IGFQAKKAGEAVALIGRTOAGNOSIALGDAOATGDSIALIGTGNVYTKHSHSAIDPDS 1903  
 2036 IGFQAKKAGEAVALIGRTOAGNOSIALGDAOATGDSIALIGTGNVYTKHSHSAIDPDS 2095  
 QY 1904 TVKADNSYVGNNOFTATQDVFVGNNITTVESNSVALGNSAISAGTHAGTOAKKS 1963  
 2096 TVKADNSYVGNNOFTATQDVFVGNNITTVESNSVALGNSAISAGTHAGTOAKKS 2155  
 QY 1964 DGTAGTTTATAGTGVGAGQOTAVGAVSAGSAGERRIQWAAAGEVSATSTDAVNSQL 2023  
 2156 DGTAGTTTATAGTGVGAGQOTAVGAVSAGSAGERRIQWAAAGEVSATSTDAVNSQL 2215  
 QY 2024 YKATOGINANNEIDRHOENKANKAGISSAMMAASPAAYIGRSMVYGGITAHNGOG 2083  
 2216 YKATOGINANNEIDRHOENKANKAGISSAMMAASPAAYIGRSMVYGGITAHNGOG 2275  
 QY 2084 AVAVGLSKLSDNGQWVEFKINGSADTQGHVGAAGAFHF 2122

Db 2276 AVAVGLSKLSDNGQWVEFKINGSADTQGHVGAAGAFHF 2314  
 RESULT 4  
 US-09-377-155-33  
 ; Sequence 33; Application: US/09377155  
 ; Patent No. 6197312  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PEAK, Ian Richard Anselm  
 ; APPLICANT: JENNINGS, Michael Paul  
 ; APPLICANT: MOXON, E. Richard  
 ; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
 ; FILE REFERENCE: 065064/0128  
 ; CURRENT APPLICATION NUMBER: US/09/377,155  
 ; CURRENT FILING DATE: 1999-08-19  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/01031  
 ; PRIOR FILING DATE: 1998-12-14  
 ; PRIOR APPLICATION NUMBER: GB 9726398.2  
 ; PRIOR FILING DATE: 1997-12-12  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 33  
 ; LENGTH: 2353  
 ; TYPE: PRT  
 ; ORGANISM: Haemophilus influenzae  
 US-09-377-155-33  
 Query Match 11.1%; Score 1185.5; DB 4; Length 2353;  
 Best Local Similarity 22.7%; Pred. No. 8.3e-61;  
 Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;  
 QY 1 MNIHYIENKATGTEMAVAEYAKSHSTGGSCATGQVGSVRLSPARIALAIVIGAT 60  
 1 MNIHYIENKATGTEMAVAEYAKSHSTGGSCATGQVGSVRLSPARIALAIVIGAT 52  
 Db 61 LNSAYAGISISEADGKGGANARGDKSLAIGDIAQ-----ALGSOSIALGDKKIYHNNS 115  
 53 -----DEDELDPVVTAAPVLSHSDKEGTEKEVENEN 87  
 QY 116 NNANIGAKAGNESIALIGDVLASGHASIALIGSDDIYLLKKEFYQOISELLPIIRGQKALN 175  
 88 WGIYFDKKG-----VLKAGATITLAKG-DNLKIKQNDTE----- 119  
 QY 176 DIYQALDNLQKRYRTHAQAASHAVGAMSYAKGHSNMFGRATAGTYSLAVGLT-AT 234  
 120 -----STNASSF-----TYSLKLDLTL 138  
 QY 235 AKAASTIAVSNNOALIGFAATANG-----GSTOVNLNGIALGFSOYVLQKDNVYNAAN 288  
 139 SVATEKLSFGANDKVDITSDANGKLAKTNGNHLN-----GLDSTLPDAVTANGVLS 193  
 QY 289 VRAVAPDDNCPINRRKATFKNGATDVFSIG-NSNGNDSIRKTIWVGASADTDVAVNA 347  
 194 SSSFTPMN-----VEKTRATYK-----DYLNAGNNIRK-----AKTAGANV- 230  
 QY 348 QLEAVEL--ANRQITFEKGDSSNN---RYEKGIGLITLTGGAQTSALTDNHIGVONG 401  
 231 ---ESVDLVSAVNNVEITGDKNTLVDVLAKEKGTTEVKFPTKTS-----YIKRK 279  
 QY 402 DGLKVLAEITLSLKAVTTENLTANEKYVYGTKRLTLDKIGFTNDMANGIDESKPYLDKOT 461  
 280 DG-----KLETKENNDITNKVT---SNTATDN---TDEGNGLVTAARAVID--- 318  
 QY 462 GIHAGGOKITLTAAGVDDAATYGOULKVNOTAESAALOTFFYKRVKDKGN-----DA- 514  
 319 AVNKAQRVATITANQONGFATVA---SGTNYTFESGDDGTASVYTDJNGNGITVAYIDAK 376  
 QY 515 -----NDSKIIYVGNKNNKPDGTQVNTL-----KLKENGVDVYT----- 548  
 377 VGDGLKFDSDKRIIVATLTALTIVTGKVAIEIAKEDDKKLLVNGADLVLTALGNSLWAKAKA 436

QY 549 EBNGTATGELNONNGITVGNSTLNDGLSVK---NTNSNKOIOWGADGITTETDINSKP 604  
 Db 437 DTDGALF-GISKDOEKAGEITVTPKAGKNLKYKOGDANFTYSLODALTGLTSTLIGCTIN 495  
 QY 605 GAGIENTRITITDGIQF--ANNITSLDANKPRITPTGINAGKELTNVOSATINPATNGO 662  
 Db 496 G-GNDAKVTINKDGLITTPAGNGGTGTNTISTYKDIGAKAKALTNVASGLR-AVDAN 553  
 QY 663 LDFMNRIST-----ANTE--KSSSAATIKOLYNSOVPPLTFAG- 698  
 Db 554 FVLNNSATDLNRHVEDAKGLLNLEKNANKOPLVTDSTAATVGLRKLGWVSTKNGT 613  
 QY 699 DDCPNVTKLIGELKKGKGTADDLTKNNIGVADSTONSILVAKLITSLD-----D 752  
 Db 614 KRESNOVKADELFTGAGAAVTYSKSENG-----KHTITVSAETKADGLEKDG 665  
 QY 753 AVNTKTLTA-SDKVTIVDSNNNTAKLONGLTFESKONTGATPA----- 793  
 Db 666 TIKLAVDNGNTDNLVGNNGGTAVTKG---FETVKTGATDADRGRVTKDATANDADKK 722  
 QY 794 -----TNSKTTIGVGLKFTDN--NGIALDGTITYTKDKVGRKOD 831  
 Db 723 VATVKDVAATINSATFVTEMLTTSIDEDNPTDNGKDALKAGDITLTFKAGKNLKYKR 782  
 QY 832 G---SLDKSKPYLDK---DKLVG-----EVEITNGIN-----AG 861  
 Db 783 GKNITFDLAKNLEVKTAKYSDTLITIGNTPTGTTATPPVNTITSTADGLNFAKETADASG 842  
 QY 862 GKAI--TGLSNTLT-ATNATGHTVLOLIVDSTDKTRAASIDVYANAGFNLKNNGDAD 918  
 Db 843 SKNVYLKGIATLTLEPSAGAKSSHVDL--NVDAKTKSNMASIEDVLRAGNIGNGNNVD 900  
 QY 919 FVSTDTVPT----- 929  
 Db 901 YATVDTVNFDTDSGTGTVTYVYQKADGKADVIGAKTSVIRKDHNGKLEFGKLDKANN 960  
 QY 930 ----- 929  
 Db 961 GATVSEDDDKDPTGTLVTAKTIVDAVKNKSGWRVTEGATFAETGATVAVNMAETVSGTS 1020  
 QY 930 ---NGNATTAKYVDGKASKVAVDVN--DGTTHLTLGADGNKNOIGVKTTLTKTDAR- 983  
 Db 1021 VNFKNKNATTAATVSKDNGINAVKYDVNVGDGLKI-----GDCKIYADTTLTTLTVGKV 1074  
 QY 984 ---GKKAINFVSNGDDKALINAKIADNLNLTAGEIR-----NTGTAPTAQ--- 1029  
 Db 1075 SVAPAGANSVN-----NNKRLVNAEGLATLNLNLTAKADKYADESESEGTQEVKAGD 1128  
 QY 1030 ---TFQV---KVKEN----- 1039  
 Db 1129 KYTFKAGKMLKYKOSKEDTYSLODITLGLTSTLIGTANGRNDGTIVINKDGLITLIAN 1188  
 QY 1040 ---GDDNDADTIVGKD-----AKTNOV 1060  
 Db 1189 GAAAGTADANGNTISVTKGISAGNKEITNVKSAKTYKDTONTADETODKEFHAUVKNA 1248  
 QY 1061 NPLUKLKGKGLDITQTKKD---TTFEGINTOSGLKAGNNTLINNG---LSIKNAGNE 1113  
 Db 1249 NEVEFGKNGATVSAKTODNNKHTVTIDV---AEAKVNGGLEKDTGKIKLKYKNDGDN 1305  
 QY 1114 QIOVGADGVKFAKVNNGVGA-----GIDGTTRITREIDEGFAGTNGS---LDRKSPHL 1163  
 Db 1306 LUTV---DARKGASVAKGEFNAVTITDATTAGTANERGVYVVGSGATATETDK----- 1358  
 QY 1164 SKDGINAGCKKTINIOSGEIAONSNDVATGKIYDLKTELENKISSHTAKTQONSLHEFSV 1223  
 Db 1359 ---KVAATV---GVAKAIINDAATFVKV-----ENDDSATI-----DDSP 1392  
 QY 1224 ADEQGNFTVSNYSYSDTSKSDVTF--AGENGITTKVNNKV---VRGIDOTKGLITP 1279  
 Db 1393 TPDGAG-----DALKAGDITLTKAGKMLKYRROCKNTITFLANDLSVKSATVSD 1441  
 QY 1280 KLVGNNGK-----GIVDSONGONT---ITGLSNTLANVTNDKGSVRTTEOGK 1326

Db 1442 KISLGTNGKNVNTSDPTKGLNFAKDKTGDNDANIHLNGIASLITLIDLINSGAR--TVLGCN 1500  
 QY 1327 IIKDEKTPRAASIVDVLASAGFNIQ-----NGEAVDFVSTYTVNPAAGNATTAKVT 1378  
 Db 1501 GLTDNEKRAASVADVLNAGMNVGRVGPASANNQVNEIDFVATYDVDFVSGKDKTSTYT 1560  
 QY 1379 YD--DTSKTSKYVYDVNDVDTTIEVADKIKGVTTTLTISG-----TGAN-KFPLSNQAT- 1430  
 Db 1561 VESKDKNGKREV-----KIGAKTSVIRKDHNGKLTGKELKLDNNNGVTY 1604  
 QY 1431 -----GDALYKASDIYVAHLNLTSLDIOPTAKASQAN-----SSAGYVADAGNVTY 1475  
 Db 1605 TETDGDENGGLYTAKAVIDAVNKAQMRVKT--GANGQNDDEFATVASCNTVTFADNGTT 1663  
 QY 1476 YDSTDNKYVQAKNDG--TVDKTEVAK-----DKLVAQA--QTPDGTILAQNNKSVIN 1524  
 Db 1664 AEVT-----KANDGSITVAKYVNVADGLKLDGDKIYADTTLTVADGKVTAPN----- 1711  
 QY 1525 KEQVNDANKKQGINEONAPVKGLEKASADKTKNAAVTVGDILNAVAQTPPLTFAGDGTITA 1584  
 Db 1712 ---NGDGK-----FVDASGLDALNKLKLSWTA--TAGEKSTGEVDPANSGQ-----E 1754  
 QY 1585 KRLGETLTIKGOTDYNKLTLDNNIGVAGTGFYKLAKDLTNLSV-----NAG-----G 1635  
 Db 1755 VKAGDKVTFKAG-----DNLKIKQSGKDEFTYSLKREKLKLTSEFEPDANGTGSSES 1805  
 QY 1636 TKIDDKGVSEFVDSGQA-----KANTPVLSANGLDIGKVIYVNGVGTK-----DT 1681  
 Db 1806 TKITKDLITTPANGKAGAGANAMNTISVTKDISAGNKAVTNNVSGLKFKFGDHTLANG 1865  
 QY 1682 DAANYOOL--NEVRNLILGNAQNDN---ADGNQVNIAD-----IKKDPNNGS- 1724  
 Db 1866 TVADFEKHYNAYKDLTNLDEKADNPPYADTATVGLDRGLGVNISDKTTGEPNOE 1925  
 QY 1725 ---SSNRTVIKAGT-----VLGG-----KGNDEKTLATG---V 1753  
 Db 1926 YNAVNRNANEFKFKSGGINVSGKTLNGTFVITFELAKGEVKSNEPTYNADGSEETNLV 1985  
 QY 1754 QVC-----VDKGNANGDLSNWWVKTQK-----DSKKALLATYNNAGQNTYL 1796  
 Db 1966 KQVDMYTSKEDIDIPATSKPMTG---KTEKYKVENGVANSAGKTEVTLTNKSG---YV 2039  
 QY 1797 ENNPAEAIERINOGIRFPHVNDGNEPVOGRNGIDSSAGSHVAIGFOAK----- 1849  
 Db 2040 TGN--OYADALAKSGFEL-----GLADAAEAKARPESKKDQLSKDKA 2081  
 QY 1850 ---ADGEAVALIGROTOAGNOSIAIGDNAQATGO-----SIAIGTVNVAKGHS 1897  
 Db 2082 ETVNANDKVFANGLMTKV---SAAVTESTDANGDKVTTTFVKTDELPLTQIY---NTD 2135  
 QY 1898 AIDDPSTVAKADNSYVGNNOPTDQTQVFGVGNNTITYESVVALGSALSASGTHAG 1957  
 Db 2136 ANGNKIVKRAKGRW-----YELNADGTASKNEVTLGNVANDNGK----- 2174  
 QY 1958 TOAKSGDGTAGTTTATGATGVKFGAQTAVGAVSG-----ASGAERIONV 2005  
 Db 2175 -VYKVTENGADKRYTNADGAADKTGSEVSNKVSDEKRIYVLDPRNOSNGGVIYDV 2233  
 QY 2006 AGEVSATSDAVANGSOLYKATQIGIAN--ATNELDHRHONENKANAGATISSAMAMASP 2062  
 Db 2234 ANGEISATSTDALINGSQLYVAVAGVTNLAGOVNNELEKVYKVRADAGTASALASQLP 2293  
 QY 2063 QATIPGRSMYTGAIATHNGGAAVAVGLSKISDNGQVYFKINGSDADPOGHGAAGVAGFHR 2122  
 Db 2294 QATMPKSMVALIAGSSYQONGIAGVSRISDNGKVIITRLSGTNSOGKTGVAAGVGYOW 2353

RESULT 5.  
 US-08-913-942-4  
 ; Sequence 4, Application US/08913942  
 ; Patent No. 6200578  
 ; GENERAL INFORMATION:





QY 1114 QIOGVADYKFAKNNGVG-----GIDGTRIRDEIGFATNGS-----LDKSKPHL 1163  
DB 1306 LITV--DARKGAVKAFENAVTTDAITTAOTANERKGVYVKGSGATATETDK----- 1358  
QY 1164 SKDGINAGKKTINIOSGEIATONSNAVYGGKIYDLKTELEKISSTAKTAONSLHEFSV 1223  
DB 1359 -----KKVATV--GDVAKAINDAATFVKV-----ENDDSATI-----DDBP 1392  
QY 1224 ADEQGNFTVSNPSSYDTSKISDVITF--AGENGITTKYKNGV--VRVGIOTKGLTTP 1279  
DB 1393 TDDGAN-----DALKAGDITLTKAGKMLKVRKRGKNTFALANDLSVKSATVSD 1441  
QY 1280 KLTGVANNCK-----GIYDSOONQNT-----ITGSNTLANTNDKSGVRTEGOK 1326  
DB 1442 KLSLGTNGKAVNTTSOTKLNFAKDSKTGDDANIHNLNGIASITLTDLLNSGAT--TNGCN 1500  
QY 1327 IIKDEKTRAASIVDVLASGFIQ-----NGEAVDFVSTYTVNPAQNAATYAKV 1378  
DB 1501 GITDNEKKRAASVYKDVNLGMVNRGVKPAASANNQVENIDFVATYDTPVSGDKDTTSVT 1560  
QY 1379 YD--DTSKTSKVYDVVNDTTEVADKIKGVTTTLTSTG-----TGAN--KPAISOAT- 1430  
DB 1561 VESKDKGKREY-----KIGAKTSYIKDHNGKLTGKELKDKANNNGVTV 1604  
QY 1431 -----GDALYKASDIVAHLNLSGDIOTAKASOAN-----SSAGYVADADGNKY 1475  
DB 1605 TETDGDGEGNGLYAKAVDAVAKAGRYKTT--GANGQNDPRTVAGSNTVFPADNGT 1663  
QY 1476 YDSTDNKYYQAKNDG--TYDKTEVAK-----DKLVAQA--QTPDGLAOMNYSVIN 1524  
DB 1664 AEVT-----KANDGSITVYKYNKVDGLKLDGKIVADTTLVYADGKYAPN----- 1711  
QY 1525 KEQVNDANKKQGINEDNAPYKLEKASDNKTKNAVYTGDLNAVAQTPLTAPGDTGTA 1584  
DB 1712 -----NDGKK-----FVDSGLADALNKLISWTA--TAGKEGTGEVDPAANSQO-----E 1754  
QY 1585 KRLGETLTIGQOTDNKLTNDNIGVAGDGTGKLAIDLTLNLSV-----NAG-----G 1635  
DB 1755 VKAGDKVTFKAG-----DNLKIKOSKDPYSLKKEKLDLTVBERKDANGGTSFS 1805  
QY 1636 TKIDDKGVSVFSSGOA-----KAMPVLSANGLDIGKVISVNGKGT-----DT 1681  
DB 1806 TKITKDGITTPANGAAGACANTANTISVTKDISAGNKAVTNVSGLKFKFGDHTLANG 1865  
QY 1682 DANNVOL--NEVRNLLGLGNAENDN--ADGNQVNIAD-----IKKDPSSGS----- 1724  
DB 1866 TVADFEKHYNAYKDLNDEKAGDNNPTVADNTAATVADLRLGLVNIADKTTGEPNOE 1925  
QY 1725 -----SNRTYIKAGT-----VLGG-----KGNNDTEKLATG-----V 1753  
DB 1926 YNAQVNRANANVKKFSGNGIIVSGKTLNGTRVITFELAKEGVKSNFETVKNADGSEINLV 1985  
QY 1754 QVG-----VDKGNANGLDSNVVKTOK-----DGSKALATYNAAGOTNYL 1796  
DB 1986 KVDMDYKSKEDIDPATSKPMTG--KTEKYKENGKRVYVANGSKTEVTLTKNGSG--YV 2039  
QY 1797 TNPAEALIDINOGIRFPIVNDGNQEPVYQNGNIDSSASGSHSAIIGFOAK----- 1849  
DB 2040 TGN--QVADIAKSGFEL-----GLADAEEAKAPAESAKDKQLSKDKA 2081  
QY 1850 -----ADGEAAVVAIGROTOAGNOSIAIGDNAQATGDO-----SAITGIVNVAKSHG 1897  
DB 2082 EYVNAHDKVFANGLNTKV--SAATVESTDANGDKVTTTFFVTDVLEPLTOIY--NTD 2135  
QY 1898 AIGDPSTVKAANSYVSGNNQTFDATTQTDVFGVNNITVETESNVALGNSAISACTHAG 1957  
DB 2136 ANGNKIYKADGKW-----YELNADGTASNKKEVTLIGVNDANGK----- 2174  
QY 1958 TOAKKSOGTAGITTTAGATGVAGGQTAVAGVSVG-----ASGAERIQNV 2005  
DB 2175 -VVKVTEENGADKWTYTNADGADKTKGEVSNDRVSTDEKHVRLDENNSNGKGVVINDV 2233

QY 2006 AAGEVATSTDAVNGSQYKATOGIAN--ATNELDRHIONENKANAGISSAMAMAMP 2062  
DB 2234 ANGELSATSTPAINSQYXANAKVNTNLAGOVNMLEGVKNVGRADAGTASALASQPL 2293  
QY 2063 QAVTGRSMVATGIATHNGGAVAVAGLSKLSNDQCMWFKINGSADTQGHVGAAGACFHF 2122  
DB 2294 QATMGKSMVAVIAGSSYOGNGLAIGVSRISDNKGVIIIRLSGTINSOGKGTGVAAGVGYOW 2353  
RESULT 6  
US-09-669-974-33  
Sequence 33, Application US/09669974  
Patent No. 6333173  
GENERAL INFORMATION:  
APPLICANT: PEAK, Ian Richard Anselm  
APPLICANT: JENNINGS, Michael Paul  
APPLICANT: MOXON, E. Richard  
TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
FILE REFERENCE: 065064/0128  
CURRENT APPLICATION NUMBER: US/09/669,974  
CURRENT FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US 09/377,155  
PRIOR FILING DATE: 1999-08-19  
PRIOR APPLICATION NUMBER: PCT/AU98/01031  
PRIOR FILING DATE: 1998-12-14  
PRIOR APPLICATION NUMBER: GB 9726398.2  
PRIOR FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 33  
LENGTH: 2353  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-669-974-33

Query Match 11.1%; Score 1185.5; DB 4; Length 2353;  
Best Local Similarity 22.7%; Pred. No. 8.3e-61;  
Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;  
QY 1 MNHITYKVFNKATGTFMAVAEYAKSHSTGGSCATGOVGSRTISFARIALAVLVGAT 60  
DB 1 MNKIRNVIMNWTQWVWVSELTTRHTK--RASATVETAVLATLFLFVQANAT----- 52  
QY 61 LNSAVAGIGISEADCGKGANAGDKSIATGDLAO-----ALGSQSIAGDNKIVHNSN 115  
DB 53 -----DEDELDPPYKRAPVLSFSDKEGTEKEVTENSN 87  
QY 116 NNANIGARASGNESIAIGDVLASGSHASIAIGSDLYLKETEVOQISELPIIRGOKALN 175  
DB 88 WGIYFDNKG-----VLKAGATILKAG--DNLKIKQNTDE----- 119  
QY 176 DIYQIADINLOKRYRTHAOGHASTRAVAGMSYAKGHFSNAPOTRATAGTYSIANGLT-AT 234  
DB 120 -----STNASSF-----TYSLKRLDLDLT 138  
QY 235 AKAAASSIIVNAQIIGPATAVG-----GSTOVNLRGIALGFGSOVLQKNDVNAAN 288  
DB 139 SVATEKLISFGANGDKVDITSDANGIKLAKTGNVHNLN-----GLDSTLPPAVNTGYLS 193  
QY 289 VRAVAPDDNOPIDNRYKATFRKNGATDVESIG--NSNGNDSIRKTIINVGASADTDAAVNA 347  
DB 194 SSSFTPPND--VEKTRATVYK-----DVLAQMNKIG-----AKTAGSNV- 230  
QY 348 QLKEAVRL--ANKQITTFGDDSN-----RVKGLGKLTITTTGGAQTSALDHNIGVQNG 401  
DB 231 ---ESVDLVSANVNEFTTGDKNTLDVLTAKENGKTEVEFTPKTS-----VIKEK 279  
QY 402 DGLKQVLAETLSLKMVTENLTANKEVTVGKTRTLTDKISGTFDMDMGIDESKPYLDKDT 461  
DB 280 DG-----KLTGKENNDTNKVT--SNFATDN--TDEGNGLYAKAVID-- 318  
QY 462 GIHAGOKITKLTLAGVDDAATVQOLKKNQVQTAESALQTFVTKVNDKNGN-----DA- 514

Db 319 AVNAGRRVKTTPANGONGFATVA--SGTNVFESGDTTASVTKDFONGITVKYDAK 376  
 QY 515 -----NSKITTVCNNKPKDGOVNTL-----KLKGEADVTT-----548  
 Db 377 VGDLKEDSKKLTAVDTTALTJTGKVAELAKEDDKKRLVAGDITLALGNSLMKAKAA 436  
 QY 549 ETNGTVEFGLNONNGTLVGNSTLNNDSLVS-----NTNSKJOIOVADGILTFEDISNSKP 604  
 Db 437 DTGDALE-GISKODEVAGETVITFKAGKNLAKODGANFTYSLDALTGLTSLTGSTTN 495  
 QY 605 GAGIENTRITRDIGF--ANNTEGLDANKPRLPTGINAGKELTVOSAINPATNGQ 662  
 Db 496 G-GNDATVINKDGLTTPANGNGTGTNTISTVKDGIKKNKAITVVASGLR-AYODAN 553  
 QY 663 LDFPNRST-----ANTE-----KSSNAITIKDLYNSOVPPLFAG- 658  
 Db 554 FDLVNSATDLNRHVEDAYKGLNLNEMNANKOPLVDTSTAAATVGDRLKLGWVSTKNGT 613  
 QY 699 DTGNVTKKIGELIKVKGKTTADLTFRNNIGVADSTDSLTVKLAKTSLD-----D 752  
 Db 614 KEESNOVKODEVLTFTAGAAITVTSKENG-----KHTTVSAETKADCGLEKOD 665  
 QY 753 AVNTKTTLTA--SDKVTYVDSGNNTAKLQNGDLFSKONTGATPA-----793  
 Db 666 TIKLVKVNONTDNLVTVGNNGTAVTKG---FETVKGATDADGKVTYKDATANDADK 722  
 QY 794 -----TNSKTIGVDLKFEDN---NGIALDGTYYITRKVGFPAKOD 831  
 Db 723 VAVKVDATALNSAATEVKTENLTTSIDENPTEGKODALKAGDITLFRAGKNLKYRD 782  
 QY 832 G---SLDKSPYLDK-----DKLVG-----EVELTNGIN-----AG 861  
 Db 783 GKNTTFDLAKLEVKTAKVSDTLTIGNTPTGTTAPKVNITSTADGLNFAKETADAG 842  
 QY 862 GKAI--TGLSNTLTD--ATNATGHTVTOIGVDSPTKRAISIGVLAGENLKNNGDARD 918  
 Db 843 SKNYLKGIAITLTPESAGAKSSHVL--NVDAFKKSNASIEVYLAGANNIGONGNVD 900  
 QY 919 FVSTYDVTDF-----929  
 Db 901 YVATYDVNFTDSTGTTVTVTKADKAGADVIGAKTSTYKDHNGKLTGKDKLDANN 960  
 QY 930 -----929  
 Db 961 GATVSEDDGKDTGLYTAKTVIDAVNKSGRVTGEGATETGATAVNAGNAETVTSCTS 1020  
 QY 930 -----NGNATTAKVYDGKASKAVADVNV--DGTTHLTGADGNKQIOVKTTLTKTDKAD- 983  
 Db 1021 VNFKNAGNATTAIVSKDGNINIVKDVAVGDLKI-----GDDKKIYADTTTLTVTGKV 1074  
 QY 984 -----GDKAINEVNSGDDKALINAKDIADINLTAGEIR-----NTKGTATALO--- 1029  
 Db 1075 SVPAAGANSVN-----NKKILVNAEGLATLNLMTAKADKYADGSEGETDEVKAGD 1128  
 QY 1030 -----TEOV---KVKEN-----1039  
 Db 1129 KVTFRAGKNLVKOSKEDFTYSLDITLGLTSLTGSTANGRNDGTIVINKDGLITLAN 1188  
 QY 1040 -----GDDNDADITLVGKD-----AKTNOV 1060  
 Db 1189 GAAGTOSANGNTISVTIKDISAGNKEITVKSALKTYPKQONTADETOKEPFAAANKA 1248  
 QY 1061 NTLKLGKNGIDJOTNKDG---TVTFGINFOSGLKAGNNLTNNNG---LSIKTAGNE 1113  
 Db 1249 NEVEFVNGKATVSAKTDNNKHTVTIDV--AAKVGDDGLEKTDKIKLKDNTGNN 1305  
 QY 1114 OIOVADGVKRAKVNNGVYGA-----GIGGTRITRDEIGFACTNS-----LDKSPHL 1163  
 Db 1306 LITV--DATKASVAKGEFNAVTTDATTAOSTNANENGKAYVKSNGATATETDK----- 1358  
 QY 1164 SKDINAGKRTINOSGEIAONSNDVATGKIYDLKTELENKISSTAKTAONSIEFSV 1223

Db 1359 -----KKVATV--GDVAKAINDAATPVV-----ENDDSATI-----DDSP 1392  
 QY 1224 ADOGNNTFVSNPSYSDTSKSDVTF--AGENGITTKVKNV---VRGIDQTFGLTTP 1279  
 Db 1353 TDDGAN-----DALKAGDTTLTKAGKNLKYKRRGKNITRALANDLSVKASATVSD 1441  
 QY 1280 KLTVGNMNGK-----GLVIDSONQONT-----ITGLSNLAVNTDNGSVRTTEOGR 1326  
 Db 1442 KLSLGTNGKNVNTISPTKLNLFPAKDSKTGDDANIHNLNGIASLTTLTLNSGAF--TNLGN 1500  
 QY 1327 ITRKEDKTRASIVDLSAGFNLCG-----NGEAVDSTYDTVFADGNATTAKYT 1378  
 Db 1501 GLTDEKKAASVQVNLNAGWVNRGVKPPASANNQVENIDFVATYDVVFSGDKDTSTYT 1560  
 QY 1379 YD--DTSKTSKVYDVVNDOTTEYKDKLGVKTTTLTSTG---TGAN-KFALSNOAF- 1430  
 Db 1561 VESKDNKRRTEV-----KIGAKTSYIKDHNGKLTFLGKELKANNNGVTV 1604  
 QY 1431 -----GDALVKASDIVAHLNLTLSGDIOTPAKASQAN-----SSAGYDADGNKYI 1475  
 Db 1605 TETDGDENGGLVTAKAVIDAVNKAAGWRYKT--GANGQNDDEATVASTNTFFADGNGT 1663  
 QY 1476 YDSTDKRYVQAKNDG--TYDKTEYAK-----DKLVQA--QPPDGTLAOMNYSVIN 1524  
 Db 1664 AAVT-----KANDGSITVKNYKVDGLKLDODKIVADTVLTVAADKVTAPN----- 1711  
 QY 1525 KEQVNDANKKQGINEDNAVYKGLERASDNKTKNAAVTYGDLNANAQTLTFRAGDTGTA 1584  
 Db 1712 ---NDGKR-----EVDASGLADLANKLSMTA--TGAKBGTGEVDNPSAGO---E 1754  
 QY 1585 KRLGELTTIKGGOTDFNKLNNIGVAGTDFVYLAIDLNLNSV-----NAG---G 1635  
 Db 1755 VKAGDVTYFRAG-----DNLIKQSGKPTTSLKKEILDLSVEKDKANGGTGSS 1805  
 QY 1636 TKIDKGVSEFVSSGOA-----KANTPVLSANGLDGKRVISVNGKTK-----DT 1681  
 Db 1806 TKITKGLTTPANGAGAGANTANTISTVKDISAGNKAIVNVVSGLTKPFDDGHTIANG 1865  
 QY 1682 DANNVOOL--NEYRNLIGLGNAGNDN--ADGNQVNIAD-----IKDPNNGS--- 1724  
 Db 1866 TVADFEKHYDNAKDLTNDKEDGANNPTVADNTAATVADLGLGVLNADTTGEPNOE 1925  
 QY 1725 -----SSNRTVYKAGT-----VLG-----KGNNDTEKLTAGC---V 1753  
 Db 1926 YAKOVNANNEVFKSGNGINWSKTLNGRVITFELAKEBYKSNFTYKANADGSTNTLV 1985  
 QY 1754 QVG-----VDKDGANGLDSNVYKTKOR-----DGSKALLATYNAAGQNTYL 1796  
 Db 1986 KVGDMYYSKEDIDPATSKPMTG---KTERKYKENGKVSANGSKTEVTLTKNGSG---YV 2039  
 QY 1797 TNPAPALDRINEOGIRFHFVNDGNQEPVYQGRNGIDSSASGKHSVALIGFOAK----- 1849  
 Db 2040 TGN--GVAPAIKSGFEL-----GLADAALAEKAPAESAKKOLSKDKA 2081  
 QY 1850 -----ADGEAAVAIGROTOAGNOSIALGDNAOATGDO-----STAIQTGNVYVAKHSG 1897  
 Db 2082 ETYNAHDKRFANGLMTKY--SAATVESTDANGKVTTPKTYDVELLTQIY---NTD 2135  
 QY 1898 AIGDPSTVRKADNSYSVGNNOFTDATQTFVFGVNNITYESNSVALGSNSAISAGTHAG 1957  
 Db 2136 ANGNKIYKRAKDW-----YELNADGTAASKVEYTLGNVDANGK----- 2174  
 QY 1958 TOAKKSDGTAGTTTGTAGATGYKGFAGQRAVGVSVG-----ASGAERIONV 2005  
 Db 2175 -VVKVTENGADKWTYNMADGADTKGEVSNKVSDEKHHVRLDENNOSNKGAVYIDV 2233  
 QY 2006 AGEVASTIDAVNGSOLKATOGIAN--ATNELDRHIRENENKANAGISSAMMASXP 2062  
 Db 2234 ANGEISATSIDALNGSOLTAIVAKGVYNNLAGOVNLEBKVKNYKGRADDACTAALASQLP 2293  
 QY 2063 QAVIPGRSVYTGIAITHNGOGAAVAVLSKLSDNGQWVERKINGSADTOGHVAAVAGAFHF 2122  
 Db 2294 QATMPKSMVAIAGSSYOGNGIAIVSIRISDNGKVIIRLSGTTNSOGKTVGAAGVGYOW 2353

## RESULT 7

US-09-268-347-47

Sequence 47, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860

CURRENT APPLICATION NUMBER: US/09/268,347

NUMBER OF SEQ ID NOS: 54

SOFTWARE: Patentlin Ver. 2.0

SEQ ID NO 47

LENGTH: 2354

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-268-347-47

Query Match 11.0%; Score 1177; DB 4; Length 2354;  
 Best Local Similarity 23.0%; Pred. No. 2.6e-60;  
 Matches 612; Conservative 327; Mismatches 876; Indels 846; Gaps 114;

1 MNHIKVFENKATGEMVAEYAKSHSTGGSCATGOVSVTLTSPARIALAVLIGAT 60  
 1 MKRIEIVINWMTQIMVYVSELRTHTK--RASATVETAVLATLLEFATVQANAT----- 52  
 61 LMSAVAGIGISEADGGCGANARBDKSLAIGDIAO-----ALGSQSLAIGDNKIVHNSN 115  
 53 -----DEDELDPVVRTPAVLSFHSDKRGTEKEVTENSN 87  
 116 NNANIGAKAGNESIALIGDVLASGHASIALGSDLLYKKEVQOISELLPIRROKALN 175  
 88 WGIYEDNKG-----VLKAGATITLKAG-DNLIKIKONTDE----- 119  
 176 DIYOLADTLMQYRTHAOGHASTAVGANSYAKGHSNAGTRATAREGYSILAVGLT-AT 234  
 120 -----STNASSP-----TYSLKDLTLDTL 138  
 235 AKAASSIAVGSNAOAIIGPATAVG-----GSTOVNLNGIALGSGOYLKONDVYNAAN 288  
 139 SVATEKLSFGANGDKVKITSDANGKLAKTGTNGNVLN-----GLDSTLPDAVTGTGVL 193  
 289 VARAYADDNQPIDNRKATFKENGATDVSIG-NSNGNDSIRKKIINVGASADTDVAVNA 347  
 194 SSETFPND--VEKTRATVYK-----DVLNAGNNIKG-----AKTAGNV- 230  
 348 OLKEAVRL--ANROITFKGDDSN--RVEKGLKTLITITGAQTSALTTHNIGVONG 401  
 231 -----ESVDLSAVANNVEFLITGDKNTLDVLTAKENGKTEFEKTPKTS-----VIREK 279  
 402 DGLKVALAETLSLKAVTTEENLTANEKVYVGTKRLTTDKIGTNDMNGIDESKPLDKDT 461  
 280 DG-----KLETKENNDNRKVT--SNRATDN--TDEGNGLVYTKAAVID-- 318  
 462 GIHAGGOKITKLACVVDADATYGGKVKVNOTAASALOTFTVKKVKKGN--DA- 514  
 319 AVNKAAGRWKTTTANGQNDGDFATVA--SGTNVTFESGDTTASVTKDINGNGLTYKYIDAK 376  
 515 -----NDSKLITVGNKNNKPDGTQVNTL-----KLKENGVDVYTT----- 548  
 377 VGGGLKFDSDKKIVADTTLATLTGKGVAAELAKEDDKKLVNAGDLVTLALGNLSWKAKAEA 436  
 549 ETNGVTVEGLNDNNGLTGVNSLINDGLSVK--NTSNKQIOYGAGDITTTDISNSKP 604  
 437 DIFGALF--GISKDOEAKGEVTFKAGKNLYKODGANTFYSLODAPLITLIGGTIN 495  
 605 GAGIENTTITTRDIGF--ANNUGSLDANKPRLTPTGINAGKELTNVOSATINPATNGQ 662  
 496 G-GNDKATVINDGGLTITPAGNGGTGTNTISVTKDGIKAGKAKAITNVAAGLR-AVDAN 553

663 LDFMNRIST-----ANTE-----KSGSAAITIKDLINISOVPLTFAG- 698  
 554 EDVLLNSATDLNRHVEDAYKGLINLNKNNANKOPLYDSTAATVAGDLRKIGWVVSSTKNG 613  
 699 DTGBNVTKKIGELIKVYGGTTFADDLTKNNIGVAASTDSLVKLAKTSLD-----D 752  
 614 KEESNOVKQADEVLEFGAGAAVYTSKSENG-----KHTITVSAEIKADCGLEKDD 665  
 753 AVNTKTLTA--SDKVTVDSGNNTAKLQNGDLTFESKONTATPA----- 793  
 666 TIRKAVNONDNTLVYNGNNTAVTKG--FEVYKGTADADRGVYVATANDADAK 722  
 794 -----TNSKTIGVGLKFTDN--NCIALDGTYYTKKGVGAKOD 831  
 723 VAVKDVATATINSATPEVTEENLTTSIDEDNPTDNGKDALAGDILTFKAGNKLKVRBD 782  
 832 G--SLDKSRPYLDK-----DKLVG-----EVELTINGIN-----AG 861  
 783 GKNTTFDLAKNLEVKYAKVSDTLTIGGNTPTGTTATPKVNTISLADLFAKETADAG 842  
 862 GKAI--TGLSNTLTD--ATNATGHTVLGIIVDSTDKTRASIGDVNAGFNKNGDAND 918  
 843 SKNYLKGIAITTLLEPAGAKSSHVDL--NVDAATKKSNAASIEDVLRAGNNGNNGND 900  
 919 FVSTYDVTDFI----- 929  
 901 YVATYDVNFTDSTGTTVYVTKADKADYKIGAKTSVIRKDHNGKLTFGDLKLDAN 960  
 930 ----- 929  
 961 GATVSEDDGKDTGILYTAKTAVIDAIVKSGWRVTEGATFETGATAVMAGNAETVSGTS 1020  
 930 -----NGNATTAQVYDGRKASRVADVNV-DGTTIHLTGADGNKNOIGVKTTLTKTDAR- 983  
 1021 VFNKGNATTFATVYKDGNGNLNVKVDVWVGDLKI-----GDPKKIYADTTLTVGKV 1074  
 984 ---GKRAINEVSNGDCKALINAKDLADNLTLAGETR-----NFKGTATALO--- 1029  
 1075 SVFAGANSVN-----NNKLVNAGLATALNLSWTAKADYKADGESBETDQEVKAGD 1128  
 1030 ---TFQV---KKVEN----- 1039  
 1129 KVTFRAGNKLKVKOSEDFYTSLODTLTGLSTLTLGSTANGRNDGTIVIKKDLITILAN 1188  
 1040 ---GDDNDADITLYKGD-----AKTNOV 1060  
 1189 GAAGTASNSNTISVTKDGSAGKELTNVKSALTKYKDTONTADETOKEFHAAYKNA 1248  
 1061 NTLKIKRNGLDIOTNKDG--TVTFGINTOSGLAKAGNNTLNNNG--LSIKNTRAGNE 1113  
 1249 NEVEFVGKNGATVSAKTDNNNGKHTVYIDV--AAKVGDDLEKPTDKIRIKLVNDTDGN 1305  
 1114 QIOVGADGVKFAKNNNGVYGA-----GIDGTTITDEIGEFAGTNGS-----LDKSPHL 1163  
 1306 LFTV--DFTKASVAKGEFNAVTTDATTAOSTNANEGKAYVAGSKNATATEIDK----- 1358  
 1359 -----KKVATV--GDVAKALINDAATFVKV-----ENDSARTI-----DDSP 1392  
 1224 ADEGNNFTVSNPYSSTDSKTSVYTF--AGEKNTITRKVKGV--VRVIGDITKGLTTP 1279  
 1393 TDDGAN-----DALAGDTLLKLGKKNLKVBRDKNITFALANDLSVKSATVSD 1441  
 1280 KLVGNNGN-----KGIYI--DSONGONT--ITGLSNTILANVNDKGSVYTFEOK 1326  
 1442 KLSGTNNKNVNTSDTKGLFAKDSKTGDANIHNLGISTLTLNSGAT--TNLGN 1300  
 1327 IIKDEKTRASIVDVSAGFNLOG-----NGEAVDFVSTYDVNFADGNATTAQV 1378  
 1561 GITNEKRRASVADVLNAGNNGVYKAPASANNVEYENIDFATYDVTDFVSGKDDTTSVT 1560  
 1379 YD--DTSKTSKRVYVNVNDTITIEVKDKKLGKVTTLTSTG---TGAN--KFAISNOAT- 1430



QY 822 KDVFAKODS---SLDKSRPYLDK-----DKLVG-----EVEITTING 859  
 Db 771 AGMLKRRKROCKNITPDLAKLEVKAKVSDTLTIGNTPTGTTATPKVNIITSTADGLN 830  
 QY 860 -----AGKAI--TGLSNTLTD--ATNATGHTVTOIGIVSDTKTRAASIGDVNAGF 908  
 Db 831 FAKETADASGSKNVYLKGIATTLTEPSAGAKSSHVL--NVDATKKSNAISIDVLRAGM 888  
 QY 909 NLKNSGAKOPEVSTYDVF----- 929  
 Db 889 NIQGNMNDVATVATVDFNEFTDSTGTTVTVYQKADGKADYIGAKTSVIRKDHNGKLF 948  
 QY 930 ----- 929  
 Db 949 TGRDLKANDNATYSEDDGKDTGGLVTAKTVIDAIVKSGWRVTEGATATETATAVNAG 1008  
 QY 930 -----NGNATTAKVYTDGKASKVAVDVNV--DGTTHLTLGADGNKNOIGVKT 974  
 Db 1009 NAETVTSSTVNFKNATATVSKDNGNINVKYDVVGDGLKI-----GDDKTIIVDT 1062  
 QY 975 TLLTKTKAK-----GDALNFSVNSGDDKALINAKDLADLNLAGEIRNTKGTADTALQ 1029  
 Db 1063 TLLTVTGKVSVPAGANSV-----NNKRLVNAEGLATLNNLSWTAKADK--YADGESE 1115  
 QY 1030 TFOYKVKVKGDDDDADDTTVGKDAKTNO-----VNTLKL-----KGNG 1070  
 Db 1116 GETDQEVKA-----GDKVTFKAGKNLKVKQSEKDFYSLDITLGLTSLTGGTANGRD 1170  
 QY 1071 LDIOTNKG--TVYFINGTOSGLKAGNNTLN--NGLS-----IKNT 1109  
 Db 1171 TGTVINKDGLTITLANGAAAGTASNGNNTISYTKDISAGNKEITVKSALTKYKDTQWT 1230  
 QY 1110 AGNEOIQVGA-----DGVYFAKVNNGVYAGIDG-----TTRIT----- 1143  
 Db 1231 AGATOPANATNAEYAKODLVDLTKPATGAAGNGADAKAPDTTAATVGLDLKGLVLSAKKT 1290  
 QY 1144 -----RDELIGFAGTNGS-----LDKSKPH-----LSKDG-- 1167  
 Db 1291 ADETOKEFHAAVKANANEVEFGKNGATYSAKTDNNGKHITVIDAEKVGDLLEDITDG 1350  
 QY 1168 -----INAGCKITINIOSGELTAONSNDATVG-----GKIY-----DL 1199  
 Db 1351 KIRKYDNTDGNMLTVYDATKASVAKGEBNATVTTDATTACGTNANERKSVYVKGSGNGAT 1410  
 QY 1200 KTELENNISSTAKTAAONSLEHFS--VADEQGNFETVSNFYS--SYDTSKTSYDITF--AG 1253  
 Db 1411 ATETDKKKATVGDVAKAINDAATFKVENDDSATIDSDPTDGDANDALKAGDTLTLKAG 1470  
 QY 1254 ENGITTIVKNGV--VRVIGDQTKGLTTPKLTIVGNNNGK-----GIVIDSONQON 1300  
 Db 1471 KNLKVRKRDGNITFALANDLSVKSATVSDKLSGTNGKNVNITSDTKGLNFKAKDSKTGDD 1530  
 QY 1301 T-----TGLSSTLANVYNDGVSVTTOGKAIKDEKTRAAISIVDVLSAGFNIOG----- 1351  
 Db 1531 ANIHLNGIISTLDTLLNSGAT--TNLGGNGITDNEKKRAASVYKDVILNAGMNVKGVYPASA 1589  
 QY 1352 --NGEAVDFEYSTYVNFADGNATTAITYD--DTSKTSKVVYDVVDTTIEVDKDKIG 1407  
 Db 1590 NNGVENIDFVATYDVFVSGDKDTSYVESKDNCKREY-----KIG 1633  
 QY 1408 VKTTLTSTG-----TGAN--KFAISNOAT-----GDALVKSADIVANLTLSGDIOT 1453  
 Db 1634 AKTSVIRKDHNGKLTGKELKLDANNNGVTEITDGKDEGNGLVTAKAVIDAIVKAGMRVKT 1693  
 QY 1454 AKGASQAN-----SSAGYVADAGNKVITYDSTDKYIYAKKDG--TYDKTEYAK----- 1500  
 Db 1694 T--GANGONDDPATVAGTIVTFADGNGTAEVY-----KADGSIYVYKVVAVDGLKL 1746  
 QY 1501 --DKLVAQA--QTPDGLTLOMNVKSVINKEQVNDANKKOGINEDNAFVFKGLEKASDNK 1555  
 Db 1747 DGRKIVADTIVTVLVADGKGTAPN-----NODGKR-----FYDASGLDALANK 1788

QY 1556 TKNAATVGDLANVAVOTPLTFAGDGTAKKIGETLITKGGOTDNNKLTNNIGVAGTD 1615  
 Db 1789 LSWTA--TAKREGTGEVDPANASAGQ-----EVKAGDKVTEFAAG-----DMLKIKOSGK 1834  
 QY 1616 GFTVRLAKDLTNLNSV-----NAG-----GTRIDDGVSFVDSGQA-----KANTPVLSA 1661  
 Db 1835 DFTYSLKFKLKPULTEVEFADNGGTGSESTKTKTKGTLTPPANGAAGAANTANTISTYK 1894  
 QY 1662 KGLDGGKVIYVNGKTK-----DTDAANVOOL--NEVRNLLGLGNAGND--A 1706  
 Db 1895 DGISAGNKRAVTVVNSGLKFKFGDHTLANGTVADFEKHVDNAKDLTNDDEKADNPTVA 1954  
 QY 1707 DGNQVNIAD-----IKKDPNNGS-----SSNRVYKACT-----VLGG-- 1739  
 Db 1955 DNTATVGDRLGLGVNISADKTTGEPNOEYVNAOVNANEVKEKSGINVSGLTNGTRV 2014  
 QY 1740 -----KGNNDTEKLATG-----VOYG-----VDKGNANGDLSNWWYVTKR-- 1776  
 Db 2015 TTFELAKGEVNASNEFTYKNNADGSETNLVKYGDMYKSKEDIDPATSKPMTC--KTEKYK 2071  
 QY 1777 -----DSKKALLATYNAAGOTNLJNNPAEADIRINEGIRFFHVNDGNOEPVQ 1827  
 Db 2072 VENGVVANSKSTETVTLTKNGSG--YVTGN--QVADALIAKSGFEL----- 2113  
 QY 1828 GRNGIDSSASGKHSVAIGFOAK-----ADGEAAVALIGRTOGNGSIALGNQ 1876  
 Db 2114 ---GLADAWEAKAPRESKDKQLSKDAEYVNAIDKVPFANGLTKV---SAATVESTD 2167  
 QY 1877 ATGDO-----SLAIGTVNVAGKHSAGIDPSTVADNYSYSGNNOPTDATTQDVF 1928  
 Db 2168 ANGDKVTTTFVKTVDLDELPLTQY---NTDANGNKLVKKAADKW-----YELNAD 2213  
 QY 1929 GYGNNTITYESVALGNSAISAGTHAGTAKKSDGTAGTTTAGATGYKAGGOTAV 1968  
 Db 2214 GYASKEVTLGNVDANGK-----YKVTENCADKMYTYTNDGAADKTKGEYSN 2262  
 QY 1969 GAVSYG-----ASGAERIONVAGEVASTDANVSOLYKATOGIAN---A 2033  
 Db 2262 DKVSTDEKHVRLDPNNOSNGKGYVIDNANGELISATSTDALNSOLAANAAGVTNLAGQ 2322  
 QY 2034 TNELDRHONENKANAGISSAMAMAMPAYIPGRSWTGTGATVNGOGAVANGLSKLS 2093  
 Db 2323 VNNELEGKYNKVRKADASTASALAASQLPQATMPGKSVVALAGSYOGNGLAIGVERIS 2382  
 QY 2094 DNGQVTEFKINSADTQGHVGAAGAFHF 2122  
 Db 2383 DNGKVIIRLSTGTTNSQKGTGVAAGVGYQW 2411

RESULT 9  
 US-09-268-347-28  
 ; Sequence: 28, Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268,347  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 28  
 ; LENGTH: 1104  
 ; TYPE: PRT  
 ; ORGANISM: Haemophilus influenzae  
 US-09-268-347-28

Query Match 9.58; Score 1018.5; DB 4; Length 1104;  
 Best Local Similarity 27.84; Pred. No. 1.7e-51;  
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;  
 QY 943 KASKVAYDVNVDTGTTIHLTGADGNKNOIGVKTTLTKTKDAKGRKAINFSVNSGDD----- 997

Db 28 KCAATYAAVAVLATATLSAT-AEANNN-----TSVTNGLNAYGDTNFNTNNSIADLEKHV 81  
 QY 998 ----KALINAKD-----IADNLTLAGELR-----NTKGTADTALQTFQYKVK 1037  
 Db 82 QDAYKGLINLEKDTNKSFLVADNMTAATVGNLRLGWLVSLSKGTFRNE--KSYQVKAD 139  
 QY 1038 ENGDDDDADDTITVGNKAKTNOVNTLKLKGNGLDITQNKDGYTFPCINTQSLKAGNNT 1097  
 Db 140 E-----VLTGSGAATVS-----SSSKDGKHTTITISYTKGSFAEVKT 176  
 QY 1098 TLNNGLSIKNTAGNEOIOVGADGVKFAKYNNGVAGIGCTTRITRDEIGFAGTNGSLD 1157  
 Db 177 DATGG-----QVADRGKAKEDEN--GADV-----202  
 QY 1158 KSKPHLSKGINAGGKKTITNIOGSEIAONSNDVATGKIYDLEKTELENKISSTAKTAONS 1217  
 Db 203 -----KKAATVK--DVAKAIINDATFVKVESTDDDIENGAAKNETTDOA 245  
 QY 1218 LHESVADDEGNNFTV---SNPYSSYDTSKSDVITFAGENGITTKYKNGGVAVGIDQTK 1274  
 Db 246 L-----KAGDTLTLKAGKLNKAKLDQNGKSVTFALAKDLDTVSARVSDKLSIGKDTNK 298  
 QY 1275 -----GLTTPKLTGYNNGNGKGIYDSONGONTITGLSTLANTYNDKGSVRETEQCK 1326  
 Db 299 VDTISDANGKLAK--TGNGNG-----QNGVHLNGIASLTL---TDTITGMTQASN 345  
 QY 1337 ITKDEKTRAASIVDLVLSAGFNLOGEAVDVSVDYFNADGNATTAKYVDYDTSKTS 1386  
 Db 346 GVAVOHNHRAASVADVLNAGMNIQNGASVDFVNAVDTVDFVNGTNNVNTVTTDTAHRKT 405  
 QY 1387 KVVVD-----VNVDDTITIEVK-----DKKLG-----VTTTTLTSTGTG 1419  
 Db 406 TVRVDVYTGLPVOYVTEDEGKTVYKVDNKKYKODGSADMDKKEVNGELAKTKVLAASG 465  
 QY 1420 ANKFLASNOATGDAVLKASDIVA--HLNLTLSGDIOTAKKASQANSSAGYVADADNGKYT-- 1475  
 Db 466 QNPVKISNVAGS---TEBNDVAVSFQOLKALO-EKQVITLITASNAANAGN--DADGKATQT 520  
 QY 1476 -----YDSTDNKKYQAK--ND-----GTVDKTRKAKDKLVAQAOTPDGTLAQM 1517  
 Db 521 LNNGLNFKFSTDELINIKIENDTFTFTPKGSV--QVGEDEGKATIQNGTFTTGLVEAS 579  
 QY 1518 NVKSVINKEOVNANDKOGIEMDNAPFKGLEKASDKTKKAATVDDIADNVAQTPLFA 1577  
 Db 580 EL-----VESLNKLGKMGVVDKDS-----GELDASVETLVKS 613  
 QY 1578 GDTGTTAKLGETLTKIGGQDTNKLTDNINIGVAGTGFVFLAKDLTNSV-----1631  
 Db 614 GDKYTL--KAGENLKV--QDGTN-----FTYALKDELITGVKSVFEPKDTA 654  
 QY 1632 ---NAGGTRIDKGVSEVDSG---QAKANTPVLNANGIDLGKYSINVGKGTCTDA 1683  
 Db 655 NGSNASTRTIKRDLTITSANGANGAATDADIKIVASDGISAKNKAIVNVSGLKKEGD 714  
 QY 1684 ANVOQLNEVRML-----GLGNAGDNADGNOVNIADIKRDPNSGSSNFTVIKAG 1734  
 Db 715 ANFNLJSSANLTKQYDAYKGLTNDEKADQOTLVAD---NT-----AA 759  
 QY 1735 TVLGKGNNDTEKLATGQVGVVDKNGANDLSNV--WVKTOKDGSKALLATFYNAAGOT 1793  
 Db 760 TV-----GDLRGLGNV--ISAOKTIGELMKEVNA-----786  
 QY 1794 NYLNNPAEALDRINEGIRFFPHVNDGNOBEVVOGRNGIDS---ASGSHVAIGFOKA 1850  
 Db 787 -----QVFNANEVFKESG--NGIHVSGTKVNRREIT--FELAK 821  
 QY 1851 DGEAAVATGROTAQNGOSIATIGNAQAOTDOSIATIGNVAVAGKHSAGISGPSV--KAD 1908  
 Db 822 DENA-----IAGFYGSKALRDNMTVAIGTGNVNAEKSGAFGPNITIEDKAG 867  
 QY 1909 NSYVGNNOFTDATOTDVFVGNN-----TTV-----T 1937

Db 868 GSYAFGNDRKIT-SKNTEFLVGNCAVNAKYKANGSDVDTEVTVKDKDKETTTVPKALCAT 926  
 QY 1938 ESNVVALGSNSAISACTHAGTOAKKSDGTAGTTTGATGATGYKGFAGQAVAVASVGSAG 1997  
 Db 927 VENSIVYLGK--STATKDKGKMLKSDGTAGNTTGTAGTGVNGFAGATVAGAVSVGSAG 983  
 QY 1998 AERRIONVAAGEVASTDAVNGSOLYKATOGIANTNELDRHIONEKANAGISSAMA 2057  
 Db 984 FERRIONVAAGEISATSTALINGSLYAAKGYTN---LAQVKNVGRADAGTASALA 1039  
 QY 2058 MASPMQATYIGRSMVYGIATNNGOGAVAVAGLSKSDNGQWFEKINGSDATOGHGAAGV 2117  
 Db 1040 ASQLPQASPMGRKSMVSIAGSSVQNGLAIGVSRISDNKVIIRLSGTNSQKGTGVAG 1099  
 QY 2118 AGFHF 2122  
 Db 1100 VGYOW 1104

RESULT 10  
 US-09-268-347-34  
 ; Sequence 34, Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, "Sheena M."  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268, 347  
 ; CURRENT FILING DATE: 1999-03-16  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 34  
 ; LENGTH: 1104  
 ; TYPE: PRT  
 ; ORGANISM: Haemophilus influenzae  
 US-09-268-347-34

Query Match 9.5%; Score 1018.5; DB 4; Length 1104;  
 Best Local Similarity 27.8%; Pred. No. 1.7e-51;  
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;

Db 943 KASKVADVAVNDGTTIHLTGADGNKNOIGVTTTLTKTDKAGKAINFSVSGDD-----997  
 Db 28 KCAATYAAVAVLATATLSAT-AEANNN-----TSVTNGLNAYGDTNFNTNNSIADLEKHV 81  
 QY 998 ----KALINAKD-----IADNLTLAGELR-----NTKGTADTALQTFQYKVK 1037  
 Db 82 QDAYKGLINLEKDTNKSFLVADNMTAATVGNLRLGWLVSLSKGTFRNE--KSYQVKAD 139  
 QY 1038 ENGDDDDADDTITVGNKAKTNOVNTLKLKGNGLDITQNKDGYTFPCINTQSLKAGNNT 1097  
 Db 140 E-----VLTGSGAATVS-----SSSKDGKHTTITISYTKGSFAEVKT 176  
 QY 1098 TLNNGLSIKNTAGNEOIOVGADGVKFAKYNNGVAGIGCTTRITRDEIGFAGTNGSLD 1157  
 Db 177 DATGG-----QVADRGKAKEDEN--GADV-----202  
 QY 1158 KSKPHLSKGINAGGKKTITNIOGSEIAONSNDVATGKIYDLEKTELENKISSTAKTAONS 1217  
 Db 203 -----KKAATVK--DVAKAIINDATFVKVESTDDDIENGAAKNETTDOA 245  
 QY 1218 LHESVADDEGNNFTV---SNPYSSYDTSKSDVITFAGENGITTKYKNGGVAVGIDQTK 1274  
 Db 246 L-----KAGDTLTLKAGKLNKAKLDQNGKSVTFALAKDLDTVSARVSDKLSIGKDTNK 298  
 QY 1275 -----GLTTPKLTGYNNGNGKGIYDSONGONTITGLSTLANTYNDKGSVRETEQCK 1326  
 Db 299 VDTISDANGKLAK--TGNGNG-----QNGVHLNGIASLTL---TDTITGMTQASN 345  
 QY 1337 ITKDEKTRAASIVDLVLSAGFNLOGEAVDVSVDYFNADGNATTAKYVDYDTSKTS 1386  
 Db 346 GVAVOHNHRAASVADVLNAGMNIQNGASVDFVNAVDTVDFVNGTNNVNTVTTDTAHRKT 405

QY 1387 KVVVD-----VWDDPTTEV-----DKLGG-----YKTTTLSTGQ 1413  
 Db 406 TVRVDVTVGLPQVYVTEDEGKTVVAVDNKKYIEAKDQGSADMDKKVEENGELAKTKIVLASG 465  
 QY 1420 ANKFAALSNQATGDAALVKAASDIVA--HLNLTLSGDIQTPAKGASQANSAGVYADGNKVT-- 1475  
 Db 466 QNPVKISVAVAG---TEENDAAVSFKQOLKALQ--EKQVTLFASNAVANGGN--DADGGKATQ 520  
 QY 1476 -----YDSTDKKYQQA--ND-----GTVDKTEVAKDKLVAAQATPDGTLAQ 1517  
 Db 521 LNLGNLNEFKSTQDDELNLINIVENDVYTFPPKGSV--QVDEDEKATLQNGTKITGGLVEAS 579  
 QY 1518 NVKSVYNKEOVNDANKKQGINEDNAFVGLKEKASQDNKTKNAAVTVGDLNVAQPLEFA 1577  
 Db 580 EL-----VESLNKMGKMGVADKQDS-----GELDGASNETLWKS 613  
 QY 1578 GDVTGTTAKKIGETLLTGQGTDTNKLTDNNIGVAGTDGFTVYKLAKDLTLNLSV----- 1631  
 Db 614 GDKVTL--KAGENLKVH--QDGTN-----FTYALADELTVKSVFEKQTA 654  
 QY 1632 ---NAGGKIDDKVSVVDSSG-----QAKANTPVLSANGLDLGKVISVNGKGTGTD 1693  
 Db 655 NGSNGASTKRTIKQDLTTSANGANGAAYTDADIKIVASDGISAGNKAVKNVYSGLKKFGD 714  
 QY 1684 ANVOOLENVRLL-----GLGNAQNDNADGNQVNDLJKDPNSSSSNRIVYKAG 1734  
 Db 715 ANFNPLISSADNLTKQYDDAKYKELTNLDEKADKQTLTYAD-----NT-----AA 759  
 QY 1735 TVLGGKGNNDTEKLATGVOVGVDKDGANAGDLSNV--WKTQKDSKKALLATYNAAGOT 1793  
 Db 760 TV-----GDLRGLGWV--ISADKTTGELNKEYNA----- 786  
 QY 1794 NYLNNPAEALDRINEGIFEFYHNVNGNQEPVYQGRNGIDS---ASGKSHVAILGFQAKA 1850  
 Db 787 -----QVRANAEVKKFKSG--NGIHSKGKTVNQRRETT--FELAK 821  
 QY 1851 DGEAAVAIGRQTOGNOSIAIGDMAQATGDSIAIGTVNVAAGKHSAGIADPSTV--KAD 1908  
 Db 822 DENA-----LAFGYSKALNDNYALTQGVNVAAEKSGAFDPNPTIEDKAG 867  
 QY 1909 NSISVGNNOFTDATQDFVGVGN-----ITV-----T 1937  
 Db 868 GSYAFGNDNRIT--SKNFEVLGVNGVNNKYRANGVDVTEFVYVDKDKGKEFTVTVPKALGAT 926  
 QY 1938 ESNVAVALGNSALSAGHACTQAKSKSDGPAAGTTTATGATGCTVKGPAQTAAGVAVSAGS 1997  
 Db 927 VENSIVYGNK---STATKDKGMLKSDGTAGTGTATGATGTVGNFPAATAGAAVSAGS 983  
 QY 1998 AERRIOWNAAGEVSATSDVANSQOLYKKAQTQGIANATNELHRIHQNNKNAAGISSAMA 2057  
 Db 984 EERRIOWNAAGEISATSDAINGSQLYAAKQVTN---LAQVNVKVKRADAGASLALA 1039  
 QY 2058 MASHPQAVTPEKRSKVTCGILATHNGGCAVAVGLSKJSDNGQWFKTNGSADTPQGHVGAAG 2117  
 Db 1040 ASQPLQASPMGKSNVSIAGSSYOGONGLAIGVSRISDNGKVIIRLSGTTNSOQGTGVAAG 1099  
 QY 2118 AGFHF 2122  
 Db 1100 VGTQW 1104  
 RESULT 11  
 US-09-268-347-30  
 ; Sequence 30. Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268,347  
 ; CURRENT FILING DATE: 1999-03-16  
 ; NUMBER OF SEQ ID NOS: 54

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1004
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-09-268-347-30

```

Query\_Match 9.1%; Score 974; DB 4; Length 1004;  
Best-Local Similarity .28.2%; Pred. NO. 5.8e-49;  
Matches 337; Conservative 146; Mismatches 330; Indels 382; Gaps 45.

0Y	1073	IQTNKDGVVEGIIQTOSGLGANNETTINNNGLSIKRTXAGMEIOIYQADGVKPAKVVNNGV	11322
0Y	1073	IQTNKDGVVEGIIQTOSGLGANNETTINNNGLSIKRTXAGMEIOIYQADGVKPAKVVNNGV	11322
Db	43	LSITYOATTTGGTISTYNGLAAYST--MNPNF---NAAGN	77
0Y	1133	GAGIDGTRTRIRDEIGFAGT--NSLQSKRPHLSKQDINAGKKRTINIOSGELIAONSNAV	1191
Db	78	-----SATDLARQ---EDGAYDGLLNEKDANK-----LLVYDDKAA	113
0Y	1192	TGGKIYDKTLELEKISTAKTAKONSLSHERSVADDEGNFTVSNFYSSDYDTSKSD---	1248
Db	114	TGVNLRKLGWVLSKNGSTREKESQOYKHADEVLEBEKDGTV-----TSSSENGHT	165
0Y	1249	ITPAGENGITTKVAKGVVAVGIDQITGTLTPKLTJVN-----NGKIVIDSONG	1298
Db	166	VTFLEKDLANK-----NATVSDKLSLGANKKVDITSDTNGLEKFAKPSING	212
0Y	1299	QN-----TITGLSNTLANTNDKSGVRTTEGCKIITKEDKTRASIVDYLSAGNLOGNEA	13555
Db	213	QNGVNHGLNGLASITDIT--GTTKSAFNG--VDVONHNRKASVADVIYLAGNINIGSAS	268
0Y	1356	VDFSVYDVTYFADAGNATFTAKVYDDTSKSKVYVYVNVNDPTTIEVKKCKLGVKTTTLLS	14151
Db	269	VDPENTYDVTDFVNGLTNNVNT--TDFAHNKKT--VRVDVIGLPLQ-----	312
0Y	1416	TGTGANKFALSNCATGALYKASDIYAHLNLSGDIQTAKGASQANSAGYDADGKVI	14757
Db	313	-----VTEDEGEFV	322
0Y	1476	YDSTDNKYYOAKNDGVDTKREVAKKL-----VAAQTPDGLTQOMVVK	15300
Db	323	---KYGNEYEYEAQSGSADMOKKVENKLAKTAKKLVSANGTIPVKISVADGT---EWD	377
0Y	1521	SVINKEOVNDANKKOGINEDNAV-----KGLE-----KAA	1551
Db	378	AYSFEKJALADQOVTLSASNAAYANGSGDAGDGKGIOTLSNGLNFKFKSTDELLINIAE	437
0Y	1552	SDNKT---KNAAYVVEG-----LNVAAQTPPLTFACDGT	1583
Db	438	NDIYTFPPKKSVOYVGDDGKATITQDGAKTTTGLVEASELYDSLNKLGMKVGTGIDGTIVT	497
0Y	1564	AKKLGELTLTKGGTDTNKLTDNNIGVAGTDGFTYKLAKLDTLNNSV-----NAG	16344
Db	498	DGTHFDL--VKSQDKVTLKAGDMLKYOEBGTN--FTYALDELDTDVYSEFDPKTANGANGA	555
0Y	1655	GTKLDDKGVSVBSSGGA-----KAMTPVLSANGDLGSKVYSNNGKGGKOTDANVQOL	1689
Db	556	STKTKDGLITTPNAGAGAANTANTISTYKQISAGKAKVKNVSGLKGFGDNNEDPL	615
0Y	1660	NEVNNLLGLGNAGN-----NADGCONVNIADIKKPDNSGSSSNRTVIRAGTVLGKGNND	1744
Db	616	T-----SSADNLMFQOYDAMAYKGLTNDEKSKGQOTPAVDN---AATV-----	656
0Y	1745	TEKLTAGVOYGVKDGKDNANGDLSNV--WYTKQDQSKKALLATVYAAAGOTNLTNNPAEA	18033
Db	657	-----GDLRGEGVW--ISADKTKGELNKEYNA-----	681
0Y	1804	IDRINEGQIRFEHNDGNOEPPVQOENGNDISSASG-----HSVALGFOAKDGEAAVAIG	1859
Db	662	-----QVRANNEVKEFKSG--NGI--NWSGKTLDNGTRETTFFELANDENA-----	721
0Y	1860	ROTOGAGNOSTAIGNMACATGDSIAIGTGNVYAGKHSAGIDPSTV--KADNSYSVGNNN	1917



Db 722 -----IAFGSGSKALBDNTVAISTGNVNAEKGAGDPNYYIEDKAGSGYAFEGNDN 772  
 QY 1918 QFTDATOTDVEGVGNNT-----IVTESNSV-----AG-----SNSAIS 1951  
 Db 773 RI-----TSKNFVLGNSVNAKRDANGVNLTEKEVKGKAGKVTYVPAALGETEVENSVL 829  
 QY 1952 AGTHAGTOAK-----KSDGTAGTTTACATGTVKGAGOTAVGAVSGASGERRIONVAA 2007  
 Db 830 GMASTAKDKGKAKSDGTAGNTTATAGATGTVNGFAGATLHGAIVSGASGERRIONVAA 889  
 QY 2008 GEVASTSTDAVNSQLYKATOGIANATNELDHRHONENKANAGISSAMAMASPOAYIP 2067  
 Db 890 GEISTATSDAINGSQLYAFAKAVTN-----LAGVKNKGRKADAGTASALASOLPOKAMP 945  
 QY 2068 GRSVNTGCIATHNGOGAVANGSLKSLDNGOWEKRINSAPTOGHVGAAGVAFHE 2122  
 Db 946 GKSWSVSIAGSSYOGSGLAIGVSRISDNGKVIIRLSGTTNSOGKTGVAAAGVYOM 1000

RESULT 12  
 US-09-268-347-24  
 ; Sequence 24, Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268,347  
 ; CURRENT FILING DATE: 1999-03-16  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 24  
 ; LENGTH: 1002  
 ; TYPE: PRP  
 ; ORGANISM: Haemophilus influenzae  
 ; US-09-268-347-24

Query Match 8.88; Score 944.5; Db: 4; Length 1002;  
 Best Local Similarity 29.5%; Pred. No. 3e-47;  
 Matches 338; Conservative 134; Mismatches 342; Indels 333; Gaps 46;

QY 1125 AKVNVGVAGIDGTETRTREIGFAGTNGSLDKSRPHLSKIDGKNAGCKITINQOSEIA 1184  
 Db 40 ATVLSATVQASAGST-----GTN-----SLNVYGNKNSNFSN---A 73  
 QY 1185 QNS-----NDAYTGKGIYDLKTELENKISSTAKTAQNSIHEFSVADEQ-----GN----- 1229  
 Db 74 NNSIADLNKQNDY-----YDGLLNLNKGTDKSK-----FLVADETTAIVGNLRKL 120  
 QY 1230 NFIVSNPYSYDTS-----KTSQVITPAGEBG-----ITTKVKNGVVRVGDIDQYGLTTPKLTVG 1284  
 Db 121 GVVSTKSTKSTKEESNOVKOADEVLEEGKDGVTYVSKSENGHVTFLPALDLANKNATVS 180  
 QY 1285 NN-----NGKGIYDSQ-----NGCNTTGLSNLTANTNKGSRTEEQKII 1328  
 Db 181 DKLSLGANGKVVDTSDANGIKFAKOGTNGNGVNLNGLASTDDPRVGKTAHLTKEI 240  
 QY 1329 KDEEDTRAASIVDVLASGFLNQ-----GNGEADVFSTVDVFAAGNATTAATVYDDT 1382  
 Db 241 SDTERNRASVGDVLNMGWNRGAKKTIGGYDNDVDVSTPDTVEFAAGANANVSITDDN 300  
 QY 1383 SKTSKVYVDVNDVDTTEVK-----DKKLG-----VKTT 1411  
 Db 301 KKT-----VAVDVTGLPQVYTEDSKTVVGVNEVEYEAQODGSADMDKVEENGKLAKT 355  
 QY 1412 TLTSITGANKFALSNQATGALAKADIYA-----HLNLTSGDIQAKASQANSAGVDA 1469  
 Db 356 VKLVSANGTNPVKISNADG-----TEDDVAASEKQALKQ-----DKOVTLSASNAVAGG-SDA 410  
 QY 1470 DGNKVI-----YDSTDNKYYQAKNDG-----TVDKTK-----EVADKDLVAAQOTPDGT 1513

Db 411 DQKATQTLGNDLNFKFKSTDELLNIKAGDVTFTPKKSGVQYGD-----GKATIDG- 466  
 QY 1514 LAOMNVKSVYINKEOVYNDAK-----KQINEDNAFVGLKKAASDNKTKAAVTVGDLNVA 1570  
 Db 467 -AKTTTGLVEASELYDSLKLKMGKVGKDG-----GATD----- 501  
 QY 1571 QTPLEPAGDTGTFTAKKLGTELTIKGQDTNKLITDNIGVAVAGTGTFTYKLAKDLTNLS 1630  
 Db 502 -----GHHTDILVSGSKTLLKAG-----DNKVRKEGTFNFTVYLDDELTVGKS 545  
 QY 1631 V-----NAGTKIDKGVSV-----DSSQA-----KANTPVLASANGDLGKVTLSNGK 1676  
 Db 546 VERKEDENGANGASTKTTKDGTLTTPANDANGAATADAKIKVADSDGISAGKAKVNVVS 605  
 QY 1677 GPKDTPAANVOOLEVNRNLGNGN-----DNADGQNVNTADIKKDPNSGSSNRTYI 1731  
 Db 606 GLKFGDANFNPLT-----SSADNLTQYDINAKGLITNDESKGKQITVADNT-- 655  
 QY 1732 KAGTVLGKGNNDTEKLAGVGVGVKDKDNGANDLSNV-AY--KTOKDGRKALLATYN 1788  
 Db 656 -AATV-----GDLRGLGWISADKKTGTGESEKYSAGVR 686  
 QY 1789 AAGQTVYLTNNPAPALDRINEGCIREFHVNDGNOEPVVGNGIDSSAGSKHVAIGQA 1848  
 Db 687 NANEVFKSGANGINSGKTLIDNGTR-----ETTEL 717  
 QY 1849 KADGEAAVAIAGRTQAGNOSTIAGNOAATGOSTIAGTGNVAGKHSAGIDPSTV--K 1906  
 Db 718 AKDEN-----IAFGSGSKALRDVTVAIGVNVNAEKGAGDPNYYIEDK 763  
 QY 1907 ADSVSVGNNOFTDATQTDVEGVGN-----ITV----- 1936  
 Db 764 AGGSYAFQNDNRIT--SKNFFVLGNGVNAKYKANGVDVETVYVKDKGKETTIVYPKALG 822  
 QY 1937 -TESNSVALGSSNSAISAGTHAGTOAKSDGTAGTTTATAGTGVKGAGOTAVGAVSVA 1995  
 Db 823 ATVENSUYLGK-----STATKDKGKNLSKSDGTAGNTTATAGTGVKGAGOTAVGAVSVA 879  
 QY 1996 SGAERIONVAAGEVASTSDAVNSQLYKATOGIANATNELDHRHONENKANAGISSA 2055  
 Db 880 SGEREIRIONVAAEISATSDALNGSQLYAFAKAVTN-----LAGOVNKKGRKADAGTASA 935  
 QY 2056 MAASMPQATIPGRSVNTGCIATHNGOGAVANGSLKSLDNGOWEKRINSAPTOGHVGA 2115  
 Db 936 LAASQLPQASMSKSNVSIAGSSYOGSGLAIGVSRISDNGKVIIRLSGTTNSOGKTGVA 995

QY 2116 VGAGFHE 2122  
 Db 996 AGVGYOM 1002

RESULT 13  
 US-08-409-995-4  
 ; Sequence 4, Application US/08409995  
 ; Patent No. 5646259  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen I.  
 ; APPLICANT: St. Geme III, Joseph W.  
 ; TITLE OF INVENTION: Haemophilus Adhesion Proteins  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Flehr, Hobbach, Test, Albritten & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/409,995  
 FILING DATE: 24-MAR-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Silva, Robin M.  
 REGISTRATION NUMBER: 38,304  
 REFERENCE/DOCKET NUMBER: A-61053/REF  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELEX: 910 277299  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1912 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 US-08-409-995-4

Query Match 7.9% Score 845; DB 1; Length 1912;  
 Best Local Similarity 22.2%; Pred. No. 4.4e-41;  
 Matches 478; Conservative 254; Mismatches 646; Indels 772; Gaps 96;

QY 1 MHNIYKVFENKATGTFMAVAEYAKSHSTGGSCATGOVS-----VRLSFARIALAVL 55  
 DB 1 MKKIFVNIWVMTQIMVYVSELTRH-----TKRLNRGDPYLAITLTFATVQANMT 51  
 QY 56 VIGATLNGSAVAGIGISEADGGKGANRQKSAIGDIAQ-----ALGOSIAIGDKTI 110  
 DB 52 -----DEDELDIPVRYAPVLSHSDKEGTEKEV 81  
 QY 111 VHSNNNANIGAKASGNESIAIGDVLASGASIAIGSDLYLKEFYOQISELLPIRG 170  
 DB 82 TENSNNIGTFEDKNG-----VLKAGATTLKAG-DNLKXKQXTDE----- 118  
 QY 171 OKALNDIYQADTNILOKRYRTHAOGHASTAVGAMSAYAKGHFSNAPTRATAGTYSLAVG 230  
 DB 119 -----XTNASSF-----TYSLKDD 132  
 QY 231 LF-ATRAKKAASIAVGSNMAOIGFAATANG-----GSTQVINRKGIALGFGSOVIQKND 283  
 DB 133 LVDLTSVAETKLSFGANGKVDITSDANGLAKTNGNVHLN-----GLDSTLPDAVTN 187  
 QY 284 VNAANVRAVAPDDNPIDRRYKATFKNGATDVESIG-NSNGNDSIRKTIINVGASADTD 342  
 DB 188 TGVLSSTPTPD-----VETRAATYK-----DVLNAGNMINK-----AKTA 225  
 QY 343 AVNAVLKAEVRL--ANROITEFGDDSNRRKELGKTLITIGGAQTSALT-DHNIGVYQ 399  
 DB 226 GGNV-----ESVDLSAVYNNVEFITGDKMT-----LDVVLTAKENKKTTEVEKFTPTSVIK 276  
 QY 400 NODGLKVLQAEITLSIKVNTTENLANEKYVYGTRLTDTKIGTFNDMANGIDESPYIDK 459  
 DB 277 EKDG-----KLETGENNDJNKVT-----SNTATDN-----DDEONGGLVTAKAVID- 317  
 QY 460 DTGICAGGKIKTLRAGVYDDAATYGOLKKVNOTAESALOTFEYTKVYDKNGN-----D 513  
 DB 318 --AVNKAAGRVTITANONGNDFATYA--SGTNVTFESGDDTTSASVTYDTONGNGLTVYKD 373  
 QY 514 A-----NDSKIITVGKNNKPGDQVNTL-----KLKGENGVDTT----- 548  
 DB 374 AKVGDGLKFDSDKILVADTALTAVTGKVAELAKEDDKKLVNAGDVLTAIGNLSMKAKA 433  
 QY 549 --ETNGTYFGLNONGNLVNGSTLANOGLSYK-----NTNSKQIQOVGADGTTTFIDISN 602  
 DB 434 EADDTGALG-EISKDOEKYKAGETVTFKAGKNNLKYKQDGANFTYSQDALTGLTISTLIGT 492  
 QY 603 KRGAGIENTRTTRDGIGF--ANNNGSLDANKPRLTPTGINAGKELINVOASAIIPATNG 660  
 DB 493 TNG-GNDKATVYINKDGLITTPAGNGTGITNTISTYTKGIRAKGNKAITNVASGLC-AIDD 550  
 QY 661 GOLDFMNLST-----ANTE--KSGSATIKDLNLSQVPLTFA 697

DB 551 ANFDVLNNSATDNLRRHVEDAVKGLLNENKMANQPLVDTSTAATVADLKLGMVYSTKN 610  
 QY 698 S-DTGPVNTKKGKGLILVYKGGKTTADDTLTKNNICVAVSDNSLYKLAITLSDL----- 751  
 DB 611 GTRKEESNOVQADEVLTTGGAAATVTSKSENG-----KHTTVSAETKADCGLEKD 662  
 QY 752 -DAVNTKTLFA-SDKVTVDSGNNTAKLQNGDLTFESKONTGATPA----- 793  
 DB 663 EDTIKLKVNDQNTNVLTVGNGNGTAVYKGG--FEYTKGATDADRKTYVKNATANDAD 719  
 QY 794 -----TNSKTIQVGLKFTDN--NGIALDGTYYITKDKYGFAP 829  
 DB 720 KKVATVQVATAINSATFVYKTEMLTTSIDEDNPTDNGKDDALKAGDTLFFKAGKMLKYK 779  
 QY 830 QDG-----SLDKSPYLDK-----DKLYG-----EVEITNNGIN----- 859  
 DB 780 RDGKNITFDLAKNLEVTAHVAVSDTLTIGNTPPGGTTATPKVNTSTADGLNFAKETADA 839  
 QY 860 AGKRAI--TGLSNTLTD-ATNATTGHVYQGIYDSTDKTFRASIGVILNFMKNNNGDA 916  
 DB 840 SGRNNVILKGIATTLTPRSAGAKSHYDL--NDATKSNAASTIEDVLRAGWNIQNGNN 897  
 QY 917 KDEVSTYDVDEFI----- 929  
 DB 898 VDVYATVDYVNEFTDSTGTTTVYVYOKADGKADVKIGAKTSYIKDHNGKLTGDKDKDA 957  
 QY 930 ----- 929  
 DB 958 NGATVSEDDGKDTGLVYAKTVIDAVNKSQWRYTEGATETGATVAVNAGNAETVTSQ 1017  
 QY 930 -----NGNATTAKVTVYDGKASKAVYDVNV-DGTTIHLTGADGNKNOIGVKTTLTKTDA 982  
 DB 1018 TSVNFKGNNTATVAVSNDGNINVKYDVNVGDGLKI--GDOKIYADTTTLTVTGG 1071  
 QY 983 K--GDKALINFSVNGDDKALINAKDIADNINLTAGEIR-----NTRGTADTALO- 1029  
 DB 1072 KVSVPAGANSVN-----NNKILVNAEGLATALNNLSWTAKADKADGESEGETDEYVA 1125  
 QY 1030 -----TFQV--KKYKEN----- 1039  
 DB 1126 GDKVTFKAGKMLKYQSEKDFYSLQDTLGLTSLGTANGRNDGTVINKQGLITTL 1185  
 QY 1040 -----GDDNDADDTITVGKD-----AKTN 1058  
 DB 1186 ANGAAGTDSANGTISVTYDGISAGKELTNVKSALKTYPKQONTADETQDEKPEHAAVK 1245  
 QY 1059 QVNTLKLKGNGLDIQTNKDG--TVTFGINFQSLKAGNNTLLNNG--LSIKNTAG 1111  
 DB 1246 NANEVEFVGKAGTAVSAKTDNNGKHVTIDV--AEAKVGDGLEKPTDGRKIKLKVNTDG 1302  
 QY 1112 NEOIQVAGADVYKPAKVVANGVGA-----GIDGTRITREIOPAGTNGS--LDKSRP 1161  
 DB 1303 NNLLTV--DAIKGASVAKGGEVNAVYTDATTAOGTINANERKVVVKKSGNGFATETEK-- 1357  
 QY 1162 HLSKDGINAGKKTNTIOSEIANSNDVATGKIDYDLKLEBNKISSTPAQNSLHFE 1221  
 DB 1358 -----KKAATV--GDVAKAINDATFPYKV-----ENDSART-----DD 1389  
 QY 1222 SVADQGNFTVSNPYSYDTSKTSVITE-AGEENGITTVKNGV--VRVGIDQTKGLT 1277  
 DB 1390 SPTDDGAN-----DALKAXDTLTLKAGKNNLKYKRDGKNITFALANDLSVSATV 1438  
 QY 1278 TPKTLVGNNGK-----GIYDSQNGVNT--ITGLSMTLVNVRDKSVYTTQD 1324  
 DB 1439 GSKLSTGNGKVVNTSDTGLNFAKSKTGDANIHNLNASTLTDTLNLSAT-TNIG 1497  
 QY 1325 GKTIKDEKTRASIVYVLSAGFNLOG-----NGEAVDFVSTYDVAFADGNATPAK 1376  
 DB 1498 GNGITDNEKKAASVRYVLAAGNVRCVAPASANNQVENIDFAITDYDVDFVSGDDTIS 1557  
 QY 1377 VTYD--DTSKTSKVVYDVNVDDTTIEYKDKKLGKVTTLTSTG--TGAN-KFALSNOA 1429

Db 1558 VYTESKDNKREY-----KICAKTSVICKDHNGKLEFTEKELKDANNNGV 1601  
 QY 1430 T-----GDALVKNASDIVAHLNLSGDIQTAKGASQANSAGYVDADGNKVYDSTD 1480  
 Db 1602 TVETEDCKDEGNGLVTRAKAVIDAIVNKNAGMSVKTT--GANGNDOPAVY--ASGVNTE---- 1655  
 QY 1481 NKYYOAKNDSTVDKTEKVAADKLVQAQOTDGTIL--AOMNYSVINKYOVNDANKKOG--- 1536  
 Db 1656 -----ADGNCT--TAEVTK-----ANDGSITVYVYK-----VADGKLKLDGDKI 1692  
 QY 1537 INEDNAVFKGLEKASDN-----KTKNAATVGDILNAVAQTPLEFAGDGT----- 1582  
 Db 1693 VADTVTVLVAADKYTANNNKXGKFXDASGLAGLKLST--ATACKEGTGEVDANSA 1750  
 QY 1583 -TAKLGETLTKGGQTDINKLTDNNIGVAVGDFTVKLAKDLTLMNSV-----NAG-- 1634  
 Db 1751 GQEVKAGDVKYTFKAG-----DNLKIKOSKXKDFYSLKELKDLTSEFKDANGTG 1801  
 QY 1635 --GKIDDKVSPVDSGQA-----KANTVYLSANGLDGKYSVNGKGT 1679  
 Db 1802 SESTKIKDKGLTTPANGAGAGANTNTISYTKDGISAGNKAVTNVVSGIK 1853

## RESULT 14

US-08-685-467-4  
 ; Sequence 4, Application US/08685467  
 ; Patent No. 6060059  
 ; GENERAL INFORMATION:  
 ; APPLICANT: St. Geme III, Joseph W.  
 ; APPLICANT: Barenkamp, Stephen J.  
 ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/685,467  
 ; FILING DATE: 22-JUL-1996  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/409,995  
 ; FILING DATE: 24-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Silva, Robin M.  
 ; REGISTRATION NUMBER: 38,304  
 ; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1912 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-685-467-4

Query Match 7.9%; Score 845; DB 3; Length 1912;  
 Best Local Similarity 22.2%; Pred. No. 4,4e-41;  
 Matches 478; Conservative 254; Mismatches 648; Indels 772; Gaps 96;

QY 1 MHYIKVLENKATGTFEMVAEYAKSHSTGGSCATGOYGS-----VRLSEFARIALAVL 55  
 Db 1 MKKIEVIMNVMTQWVAVVSELTRH-----TKRLNRGDPVLALLILRAVYQANF 51  
 QY 56 VIGATLNGSAVAGIGISEADGCKGAMARGDKSIAIGDIAQ-----ALGSOSIAGDKI 110  
 Db 52 -----DEDELDPVVRAPVLPSPHSDREGTEKEV 81  
 QY 111 VHNNSNNANIGAKAGNESIAIGDVLASGHASIAIGSDDLTKREYQOISELLPIIRG 170  
 Db 82 TERNMNGIIEFDKNG-----VLKAGAITLKAG--DNLKXKQXTDE----- 118  
 QY 171 QKALNDIYLAQNTIOLQKRRTHAAGHASTAVAGAMSAKGFHFNAGTATAGTSLAVG 230  
 Db 119 -----XTNASSF-----TYSLKAD 132  
 QY 231 LT-ATKASASIAVGSNAOAIFFAATAVG-----GSTOVNLNRGIALGFSQVLOKMD 283  
 Db 133 LLDLITVATREKLSFGANGKQVDTSDANGKLAKTNGVHLN-----GLDSTLPDAVTN 187  
 QY 284 VNAANVRAYPDDNOPIDNKRYATFNKATDYFSIG--NSNGDSIRKLIIVGASADTD 342  
 Db 188 TGVLSSTSTFPND--VEKTRATVK-----DVLNMGWNIG-----AKTA 225  
 QY 343 AVNVAOLKEAVRL--ANROITFKGDSNNRVEKGLKTLITGAQTSALT--DHNIGVQ 399  
 Db 226 GGVN-----BSVDLVSAYNNVEFTGDKN-----LDVLTAKENKTTVKTPTSVIK 276  
 QY 400 NGDGLVQLAETLTSLKMYTENTLANKRYVTKRRLTDTKIGFTNDMGIDESKPYLDK 459  
 Db 277 EKGK-----KLTGKKNNDTNKYT--STATDN--TDEGGLVYAKAVID- 317  
 QY 460 DTGIIHAGGQITKLAGVVDDAATYGLKKNQPAESLQTFYKVKVKN-----D 513  
 Db 318 --AVNAGMVRVKTITANGONGDEFAVA--SGNVPEESDGTASVTKPTNGITVYKD 373  
 QY 514 A-----NDSKITIVGNKPKDGTQVNTL-----KLKENGVDVT----- 548  
 Db 374 AKVGDGLKFDSDKIVADTALTIVTGKVAELAKEDDKKLVNAGDVLALNLSMKAKA 433  
 QY 549 --ETNGVTTFGLNQNGVLVGNSTLNDGLSVK-----NTNSKQIQVGADITFTDISN 602  
 Db 434 EADTDALE--GISKQDEVKAGEVTFKAGKLVKQDANFTSLDALTGLTSLTIGT 492  
 QY 603 KPGAGIENTRITRIGIGF--ANNTSILANKRRLPTGINGAGKRLTVQASINATNG 660  
 Db 493 TNG--GNDKATVINKDGLTTPANGGTTGTNTISVTKDGIRAKNKAITVNASGLR--AYD 550  
 QY 661 GOLDFMNRST-----ANTE--KSGSAATIKDLYNLISOVPLTPA 697  
 Db 551 ANEDVLNNSATDILNRHVEDAYKGLINLEKNANKQPLVTDSTAATVGDRLKLGWVSTYN 610  
 QY 698 G-DTGPVNTKIGELIKVKGKTTADDLTKNNIGVAVDSTNSLYKLAKTISDL----- 751  
 Db 611 GTFEESNOVQADAEVLFTGAGATVTSKSENG-----KHRTIVSAVETADCGLEND 662  
 QY 752 -DAVNTTILTA--SDKYVDSGNNFTAKIQNGDILFEKONTGATPA----- 793  
 Db 663 GDTIKIKLVNDQNDNDVLTVGNNGTAATVKG--FEYVTKGATADAGKRYVKAATANDAD 719  
 QY 794 -----TNSKITGVDLKFTDN--NGIALDGTYYITRKDYFAK 829  
 Db 720 KKVATVADVATAINSAATFKTEMLTTSIDEDNPTDGNKDALKAGDTLFFKAGKMLKYK 779  
 QY 830 QDG--SLDKSKPYLDK-----DKLAVG-----EVEITTINGIN----- 859  
 Db 780 RDGKNITFDLAKNLEVTAKVSDTLITIGNTPTGTTAPKPVNITSTADLNAKKTADA 839  
 QY 860 AGGKAI--TGSNTLD--ATNATGHTVQGLYDSTDKTRAASIGDVLNAGFNULKNNGDA 916  
 Db 840 SGSKNVYLKIGATLTPSAGAKSSHVDL--NDATIKKSNASILEDVLRAGMHIQNGNN 897  
 QY 917 KDFVSTYDVTDFI----- 929

```

      1111111111
Db 898 VDYVATYDVTNFTDSTGTTVTYQKADGKADYKIGAKTSVIRKDHNGKLTFTGDKLDA 957
      1111111111
QY 930 ----- 929
Db 958 NNGATVSEDDGKDTGLTAVTAVIDAVNKSWMRYTGEGATAETGATAVNAGNAETVTSG 1017
QY 930 -----NGNATAKYTYDASKAVADVNV-DGTTIHLTGADGNKNOGVKTTTLTKTDA 982
      1111111111
Db 1018 TSVNKNKNATTAATYVSKDGNINVKYDVAVGDKL-----GDKKIYADTTTLTVGG 1071
QY 983 K-----GDKAIFSVNSGDKALINAKDIAONLTLAGEIR-----NTKGTADTALQ- 1029
      1111111111
Db 1072 KYVSPAGANSVN-----NNKKLVNAEGLATLNNLSWTAKADKADYDESEGETDQEVKA 1125
QY 1030 -----TFQV-----KKYKEN----- 1039
      1111111111
Db 1126 GKRVTFFKAGNKLKVKQSEKDTFTYSLQDPTLGLTSTLGLGTANGRNDGTAVNKDGLTTL 1185
QY 1040 -----GDDNDADPTTYGK----- 1058
      1111111111
Db 1186 ANGAAGATDASNGNTISVTKDGISAGNKEITVRSALKTYYDQNTADETDQKEPNAVK 1245
QY 1059 QVNTLKLKAGKNGLDIQTNDG-----TVTFGINTQSGKAGNNTTLNNG-----LSIKNTAG 1111
      1111111111
Db 1246 NANEVEFGKNGATVSAKTDNNGKHTVTVIDV-----AEAKVGGLEKDTGKIKLVNDTDG 1302
QY 1112 NFOIOVAGDVAFKAVNNGVCA-----GIDGTRITRDEIGFAGTNGS-----LDKSKP 1161
      1111111111
Db 1303 NMLLVY--DATGASVANGEFMAVTTDATTAOGTANERGVYVGSNGATATEIDK--- 1357
QY 1162 HUSKDGINAGKKTINIOGSEIAONSNDAYTGKITYDKTELENKISSTAKTAONSLEH 1221
      1111111111
Db 1358 -----KKVATV--GDVAKAINDAAEFVKV-----ENDDSATI-----DD 1389
QY 1222 SVADQGNFTYNSNYSSDTSKTSIDVTTF-AGENGITTKYKGV---VRGIDQTKLT 1277
      1111111111
Db 1390 SPTDGAN-----DALKADPTLTLAKGNLKVKRQGNKLTLPALNDLSYKSAFV 1438
QY 1278 TPRLTGVNNGK-----GIYIDSONGONT---ITGLSNTLANVTNKGSRTEQ 1324
      1111111111
Db 1439 SOKLSIGTNGKNVNTSDTKGTFPAKDSKTGDDANIHNLGSLASTLTDLLNSGAT-TULG 1497
QY 1325 GKIIKDEKTRASTIVDLVLSAGFNLG-----NGEAVDFVSTYDVTNFPADGNATTAK 1376
      1111111111
Db 1498 GNGITDNEKKAASVADVILNAGMNVGVKPPASANNQVENIDFVATYDVTDFVSGDKDTTS 1557
QY 1377 VTYD--DTSKTSKYVYDVNVDDTTEVKKLGVTTLTSTG-----TGAN-KFALSQA 1429
      1111111111
Db 1558 VIVESKDNKGRTEV-----KIGAKTSVIRKDHNGKLTFTGKELKDAKANNNGV 1601
QY 1430 T-----GDALVKASDIVAHLNLSGDIQTAKASOANSAGYVADGNKVIYDSTD 1480
      1111111111
Db 1602 TYTEHDGDEGNGLYTAKAVIDAVNKAQMRVKT--GANGQDDPFTV-ASGTNTYF----- 1655
QY 1481 NKTYAKADGTYDTRKEVAKDLVAQAOTPPGTL-AQNNVKSVINKEOVNDANKKQG--- 1536
      1111111111
Db 1656 -----ADNGT---TAEVTK-----ANDGSIIVKYKVV-----VAAGLKLGDKI 1692
QY 1537 INEDNAFVKGLEKASDN-----KTKNAAYTVGDLNAAQPLTFPAGDTGT----- 1592
      1111111111
Db 1693 VADTIVLVADGKYTAPNNKGKXKXDSGLAGCLNKLSKT--ATAGEGGEVDPANSA 1750
QY 1583 -TAKKLGELTITKGGOTDNTKLTNNNIGVAVGTFVYKALKDLTNLNSV-----NAG-- 1634
      1111111111
Db 1751 GQEVKAGKVFEPKAG-----DNLKIKOSXKDFYLSLKEKLDLTSVEFKANGSTG 1801
QY 1635 --GTFIDKGVSVNDSSQA-----KANTPYLSANGLEDGKRVISNNGKGT 1679
      1111111111
Db 1802 SESTKITRDXGLTTPANGAGAAGANTANTISVTKDGISAGNKAVTNNVSGLK 1853

```

RESULT 15

```

us-09-268-347-32
: Sequence 32, Application US/09268347
: Patent No. 635182
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
: FILE REFERENCE: 1038-860
: CURRENT APPLICATION NUMBER: US/09/268,347
: NUMBER OF SEQ ID NOS: 54
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 32
: LENGTH: 1094
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-09-268-347-32

Query Match      6.5%; Score 693; DB 4; Length 1094;
Best Local Similarity 25.7%; Pred. 1.5e-32;
Matches 310; Conservative 159; Mismatches 464; Indels 274; Gaps 53;

QY 1076 NKDGTVFEGINTQSGKAGNNTTLNNGLS--IKNTAGNEQIOVAGDVAFKAVNNGVVG 1133
      1111111111
Db 2 NKIFNVIMNWTQTWAVVSELTTRAHTRKASATYATVATOLATATAEANSSASVTSRLNV 61
QY 1134 AGIDGTRITRDEIGFAGTNGSLDKSKPHLSKDGINAGKKTINIOGSEIAONSNDAYT 1192
      1111111111
Db 62 YG-DTNKFN-----AANNSIADLNK---QNDGVHGLNLNENKANKLLVDNNTAAT 111
QY 1193 GKIYDL-----KTELENKISSTAKTAONSLEHFSVADQGNFTVSNPS----- 1238
      1111111111
Db 112 VGLDRKLGWVSTKNGKENESQOVKQADVLFGSKGVQVSTSENGHATTFALAD 171
QY 1239 -SYFTSKTSOVIYTFPAGENGITTKYKNGVAVRGIDQTKLTPLTV-----GNANKGIV 1292
      1111111111
Db 172 LDMRTAVSDPLTIGGS--TT-----TGSATTPRVNVTSTAAGLNFPAKAT 215
QY 1293 IDSQNGONTI--TGLSNTLANVTNDKSVRTEQKLIKDEDTTRAASIVDLVLSAGFNQ 1350
      1111111111
Db 216 --GANGGTYHLNINASTLDDTLNLTNGVSKLDONGITADE-KKRAASQVDVLSNGMNR 272
QY 1351 G-----NGEAVDFVSTYDVTNFPADGNATTAKVYDTSKTSKYVYDVNVDDTTEVKK 1404
      1111111111
Db 273 GVKGTATSDNVDFVRYTDVEFLSGSEETTLVTVSSENGSKTKYKIGAKTSVIREKD 332
QY 1405 KLGKTTTLTSTGTC--ANKFALSQA-----TGDLVLYASDVLVALNLTLSGDIQTAKA 1457
      1111111111
Db 333 KL-----FTGANKDTNQVASNNAADPTDEGKGLVTAETVINAIVNKAQMRKITTGAN 384
QY 1458 SOA-----NSSAGYVADGN---KVIYDSTDN---KYYQAKNDG-----TVDKT 1495
      1111111111
Db 385 NQAGQFETVYSGNIVNFPADGNGTAVYTGATNGITVYKAEKAGDGLKIGNQKITADTT 444
QY 1496 K-EYAKDKLYAQAOTPDGTLAQNNVKSVINKEOVNDANKKQGINEDNAFVKGLEKASDN 1554
      1111111111
Db 445 ALFTVGGKYTA---PDAT---NGRKLIVASGLAALNK-----LSMTAKAED- 486
QY 1555 KTKMAATVGDILNAAVQPLTFPAGDGTAKKLGELTITKGGOTDNTKLTNNNIGVAVT 1614
      1111111111
Db 487 -TANG--GELD-----GTADEKEVKAQETVTFPAK-----NLKVKQDG 522
QY 1615 DGFYVLKALDLTNLNSV-----NAGCTKIDKGVSEFVSSQA--KANTPYLSANGLDL 1666
      1111111111
Db 523 ANFTYSLQDALTLGLTSTLGTGNNGAKTEINKDGLITTPANGAGANNANTISTYTKDQISA 582
QY 1667 GKVYISVNGKGTDTDAANYQOLNEVARNLL-----GLNAGNDNADNQOVNIAD-- 1715
      1111111111
Db 583 GQOSVKNVWSGLKRFGANPDPLTSSADNLTQYDDAYAGTLNDEKAGADKOTLTVADNT 642
QY 1716 -----IKKDPNSG---SSSNFTYIKAGTV--LGGGNNDTEKLANTGQVQV 1757
      1111111111
Db 643 AATVGLDRKLGWVISADKTTGELDKKEYNAQVRNANEFKRSNGINVSGLTYNGRREITF 702

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:35:52 ; Search time 51.53 seconds  
(without alignments)  
3956.949 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 NMHITKVFENKATGTFMVA.....NCSADTQGHVGAAGVAGPHF 2122

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-71:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1246.5	11.6	2059	2 D82671	surface protein XF
2	986	9.2	1588	2 A86036	probable adhesin Z
3	986	9.2	1588	2 H91188	probable adhesin E
4	728	6.8	1190	2 A82615	surface protein XF
5	670	6.3	1107	2 AC0976	probable autotrans
6	669	6.2	3705	2 AD0123	probable autotrans
7	651.5	6.1	2340	2 B71704	cell surface antiq
8	645.5	6.0	5291	2 B85547	hypothetical prote
9	634	5.9	5188	2 B85547	probable RTX famil
10	630.5	5.9	2249	2 A41477	190K surface antiq
11	622	5.8	6713	2 B89921	hypothetical prote
12	598	5.6	3013	2 AB0480	probable invasins Y
13	590.5	5.5	2271	2 F90073	hypothetical prote
14	568.5	5.3	2554	2 AB3528	extracellular seri
15	550.5	5.1	2021	2 A97859	190-KDa cell surfa
16	549.5	5.1	2535	2 AC0304	probable hemolysin
17	548	5.1	4919	2 T31105	hypothetical prote
18	546.5	5.1	2020	2 C48399	ABC-type transport
19	541	5.1	2660	2 E85822	probable invasins Z
20	539.5	5.0	2703	2 H81193	hemagglutinin/hemo
21	536.5	5.0	4152	2 T31102	filamentous hemagg
22	534	5.0	3624	2 AD0835	large repetitive p
23	533	5.0	658	2 AH0110	probable surface p
24	530	4.9	3890	2 C89921	hypothetical prote
25	529.5	4.9	1910	2 AF0394	probable adhesin h
26	527.5	4.9	5627	2 C83339	hypothetical prote
27	521	4.9	3295	2 AE0074	probable adhesin Y
28	519.5	4.9	2468	2 A83412	hypothetical prote
29	517.5	4.8	4936	2 AH2515	hypothetical prote

30	517	4.8	2481	2 D90011	FmbB protein [limp
31	517	4.8	3029	2 S76109	hypothetical prote
32	513	4.8	2514	2 F81045	hemagglutinin/hemo
33	508	4.7	1536	2 A43855	high-molecular-wet
34	508	4.7	2383	2 D64962	probable membrane
35	500.5	4.7	1651	2 JC1340	outer membrane pro
36	489.5	4.6	1477	2 B43855	high-molecular-wet
37	487	4.5	2479	2 F87386	conserved hypothet
38	483	4.5	1999	2 AB2018	hypothetical prote
39	480	4.5	1643	2 D71630	outer membrane pro
40	478.5	4.5	3591	1 S21010	filamentous hemagg
41	477.5	4.5	2055	2 T31110	extracellular matr
42	477.5	4.5	3535	2 E83641	probable hemagglut
43	476	4.4	2893	2 A64556	toxin-like outer m
44	474.5	4.4	1577	2 A35140	hemolysin A precu
45	467	4.4	1487	2 AG2560	hypothetical prote

## ALIGNMENTS

RESULT 1  
D82671  
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 20-Aug-2000  
C:Accession: D82671  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; PMID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82671  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2059 <SIM>  
A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF4338.1; GSPDB:GN  
A:Experimental source: strain 9a5c  
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accencio, M.; Alvarenga, R.  
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriro, D.M.; Carre  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
Submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; La  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Verjovski-Almeida, S.; Vettore, A.L.  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1529

Query Match 11.6%; Score 1246.5; DB 2; Length 2059;  
Best Local Similarity 24.3%; Pred. No. 1.6e-40;  
Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108;

QY 62 NGSAVYAGIGISEADG-----KGGANARGDKSTAI-----GDIQA 97  
DB 36 NAOYVINDGM---DGGCQRIYDNGSASGVRYVATQCSDEPTVYGVARRFGSGIAAD 92  
QY 98 LG-SOSIAIDNKKIVHNSNNANIGAKASGNEIAIG-GDVL-----ASGHASTAIGSDDL 151  
DB 93 QGASRNLTIGSLVY---NSGOVYDVYDKTYIRMSGVITMTNVTAGNAIAIGS--- 145  
QY 152 YLKRETVQQLISELLPIIRGOKALNDIYOLADTWLQKRRRHAGHASTAVGA--NRYANG 209  
DB 146 -----AASSADALKAKSLATKAGS 164  
QY 210 HFSNAFTRADEGTYSVLAVGTATATA-KAASIAVGSMAQAIGFAATVAVGSGTVMLNG 268

```

Db 165 ARAIALGAKASADGVDFVALGSGATAGTASSIALGLMSAVN-GAVAVGGALVTPPG 223
Qy 269 -IALGGSOYLQKNDVNAANVAIYAPDNOPIDNRYKATFKNGATFVPSIGNSGNDSI 327
Db 224 AVALGINSVA-----STGKLSGSDP-KTKTSTDSAAWKSTLAAY-SIGDVSSTNKK 275
Qy 328 RRKIIVNGASADTDVAVNAOLKEAVRLNR--QITEFGDSDNRVE-----KGLGKTL 379
Db 276 TROLSLAGTSMITDAVNAOLKVDEIASRGNLTAASGNSGNAVPGSSVDLKTNDKL 335
Qy 380 TIT-----GGAQTSALTDHINIGVONGD-----GLKVOLET--LTSLKWYTE 421
Db 336 TITKALGSDNOYFNKNDKYVTTLAVGDALINTDIALGFDVSLSTGLAITDGPVATAS 395
Qy 422 NLTAKEV-----TVGKTRLTBKIG-----FTNDMNGIDES----- 453
Db 396 GIDAGSKVISHVAGAVSET--STDVANGSOLNAVQVQASQPVTFGNGGAVKRSLSQSV 453
Qy 454 -----KPLDKOTG-IHA-----GGOKITKLTAQVVD 481
Db 454 VISGESSTAGTYSGNLKSVDPAAGRIHQLADSPKFGNVYINNGKISGVTAGTEED 513
Qy 482 AATYGLKRVNQTASALQTFYKVKYDKNGDANDSKITTYGKNKPDGTQVNTLKLGE 541
Db 514 AVNFSOLKSTISAVD--OGMTLTAAGSNG--SKVASG-----TYDLKNT 554
Qy 542 NGDVUTTEING--TYTFGLNQN--NGLTVGNSTINNGLSKYNKNSKQIQVQADGT 595
Db 555 DG-NLITISGSDNDVFNLSKDEKVDGMTSGTIVVNDGKV--GSDVALGTTGLT 608
Qy 596 FTDISNKPAGIENTRTITRDIGFANNGLSADANKPLRTPGLINAGKELTNVOSAI- 654
Db 609 ITD-----GPAVITASGIDAGSKVISHVAGAV 635
Qy 655 -----NPATNGQOLDENRULSTANTEKSGSAATIKLYLSQV--PLTFADGTGNVTK 706
Db 636 SETSTOAVNGSOL-----NAVQVQASQPVTFGMEGA-VKR 670
Qy 707 KLGEILKVGKGTATDLDLRKNIIGVADSDNSLTFVKLKT----- 747
Db 671 SLGOSVVISGESSTAGMYSGNLKSVYDEAGRIHQLADSPKFGNVYINNGKISGVT 730
Qy 748 -LSDDAVNTKTL--TASDK--VTVDGNNMTAKL-----QNGDLTFSKONTGA 790
Db 731 GTEBDAVNFQSKISFIAVDGWTLTASGNSGKVASGCTVDLKTNDLITISGSDN 790
Qy 791 TPAITN-SKITIGVDGLFTDNNGLALDGTITKD--KVFPAKODGSLDKSKPYLDKDKL 847
Db 791 DVVFNLSKDEKVDG-----TSGTIVVNDGVKVG-----SDVALGTTGLT 831
Qy 848 VGE-VETTINGINAGKAITG-----LSMTLPDATNATGHTVTOGIDVSTDKTRAASIG 901
Db 832 ITDGPVATVSGIDAGSKVISHVAGAVSETSTDAVNG-----SOLNAVQVQASQPVTFG 886
Qy 902 DVINAGFNLKNNGDANDFVSTYDVFINGMATTAKTAVYDGRKASVAVDVTGTTIHLT 961
Db 887 ---NEGAVKRSLSQSV-----VISGESSTAG-TYSGNLKSVDEAAG--RIHLQ 930
Qy 962 GADGKNNOIGVKTITLTAKDAGKRAINFVSNGDDKALINAKDLADNLTLAGELRN-T 1020
Db 931 LADSPF-FG-----NAVING-----GKISGVT 952
Qy 1021 KGTADP-ALOTFOVKVYKENGDDNDADITVIGKDAKTNOVNTLKLKNGLDIOTNKDG 1079
Db 953 AGTEEDAVANFSOLKSI-----STAVDQGTTLTASGANGSKVASG--G 993
Qy 1080 TYTFGINTOSGLKAGNNTLNNNGLSIKNTAGNEOIQVQADG--VKFAKVVNNGVAGAGID 1137
Db 994 TV-----DLKNTDGNLITISGSDNDVFNLSKDFKVDGMTS 1030
Qy 1138 GTTRITRDEITGRA-----GTNSLDSKPHLSKDGINAGGKITTINQSEIATONSNDAYT 1192

```

```

Db 1031 GTTVVNDGVKSDVALGTTGLTANGPAVATASIDAGSKVISHVAGAVSETSDAVN 1090
Qy 1139 GKITDLELEENKISSITAKTAQNSLHEFVSVADEQNNFTVSNPSSYDTSKIDVTEFA 1252
Db 1091 GSOLNAVQVQASQPVTFG-----NEGAVKRSLSQSVIS-----GESSTAGTYS 1135
Qy 1253 GENGITTYKNGVYVAVGIDOTKGLTTPKL--TVGNNNKGIVIDSQNGQNTITLSMTLA 1310
Db 1136 GGN-LKSVYDEAGRIHQLA--DSPKFGNVYINNGK-----ISCVT----- 1175
Qy 1311 NVYNDKGSVRTEDQRIIKDEDKTPRAASIVDVLASGFNLQNGEAVDVFSTYDVFNF-A 1368
Db 1176 -----AGTEEDAV--NFSOLKSISTAVDQGTTLTASGANGSKVASGVTDLKNT 1223
Qy 1369 DGNATTAVTYDDTSKTSKVYDVAVD-----DTTIEKDKKLGVKTTTLSTGTGA 1420
Db 1224 DGNLITSK-----SGDSNDVFNLSKDEKVDGMTSGTIVVND--GVKQSDVALGTTG 1275
Qy 1421 NKFALSNQATGDALVAKASDIYVAHLNTLSGDIQTAKKASQANSAGYVADADNKVITYSTD 1480
Db 1276 LTIANGPAVITASGIDAGSKVISHV-----AAGAVSETSTAVANGSOLNAVQVQASQ 1326
Qy 1481 NKYYQAKNDGVDTKTEVAKDKLVAQAOQTPDGLTAQNNKSVINKEO----- 1527
Db 1327 PVTFG-NEGAVKRS--LGQSVVISGESSTAGTYSGNLKSVDPAAGTITHLQADSPKF 1383
Qy 1528 ---VNDANKKQGI--NEONAF--YKLEKFAASDNKTKMAAVYV-----DL 1566
Db 1384 GNVYINNGKISGVTAGTEEDAVANFSOLKSISTAVDQGTTLTASGANGSKVASGVTDL 1443
Qy 1567 -NAVAQTPLETFAGDITGTTAKLGETLTK-----GGQTDINKLTDNNIGVAG 1613
Db 1444 KNTDGNLITISGSDNDVFNLSKDEKKSITVGNTOLODKKQVAVSSNVLLDSIELYITS 1503
Qy 1614 TDGF--TVKLADJYNLANSVNAAGTKIDDKVSPVDSQQAANTPVLSANGLDGLK 1669
Db 1504 HSSTSVKTLANGESVVRVTVNGDGVNID--VYVYVNDLGLSIVGASLILSGINSGH 1561
Qy 1670 VSNVNGKGTDPDANVOQLNEVRNLGLG--NAGDNDANQOV--NIADIKK-DPN-- 1721
Db 1562 KITNTAGTEDTDVAVNSQLKSVSEAVDKGWTLTASGANGSKVYSGCTVDLKTNDGLAI 1621
Qy 1722 SGSSSNRTVI-----KAGTVLGKGNNDTEKLTAGVQGVYDKDNDANG--DLSNV 1770
Db 1622 SKSGSDNDVFNLSKDFVDEVTAG--NTVNTDGVKVGSDVSLGANGFLIANGPSV 1676
Qy 1771 WKTQDKSKKALLATYTAACQTYTLNPA--EALDRINEGCIREFPHVNDGQEPVQ 1827
Db 1677 TASGFNADGK--YISHVAVGMADTDVAVNSQLQAVQSVYVKAATRYYSTNDG--TQ 1729
Qy 1828 GRNGIDSASGKSHVAIFQAKADGEAAVAIGROTOA-GNOSIAGIGNAOTGDOSTAIG 1886
Db 1720 GGNIDGDATSKALIAAGVGTQASGEGAAVGSMAASGKSTAIIGNALIASADGVALG 1789
Qy 1887 TG-----NVVAGKHSQ-----AIGDPSTVADNSYSVGNNOFTDAT----- 1923
Db 1750 DGAKDGAGASITYGKSYGVONTVGTFSVDKAGETRSISYVADKAEAMDVAVNLQOLD 1849
Qy 1924 -----QT-----DVFQV--GNNTYTES--NSVALGSNSAISAGTHAGTQ 1959
Db 1850 AVAOKSNLOTDMDRHEINIEDVEFKITKGDSASSVKMGVAMAIIGNAAVS-GTESVAL 1908
Qy 1960 AKKSDGTAGTTTGAANGTVKGFAGQTAAGVAVSAGSAGEERLIONVAGEVASTPDAVN 2019
Db 1909 GKNTVASDANAVALG-NSVADRA-----NSVSGSGSEQOYTNVAG--TADDAVN 1959
Qy 2020 GSQYKATQIATNATNELDRIHONENKANAGISSAMAMAIMPQATIPGRSMYTGIIAT 2079
Db 1960 VSOL--NOGLITAKQYTDGAVGNLRETSGVAAALATANLPAYVQCGMTSVGSSV 2016
Qy 2080 NGQGAVANGLSKLDNQOWYFKINGSADTOGHVAGAGAFHF 2122
Db 2017 QGQSAIAGVAVSSESGHWFKFSGSANTRSHVGVAGAGVQW 2059

```



```

RESULT      2
AB6036
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001
C:Accession: AB6036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallantia, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5460; MUID:21074935; PMID:11206551
A:Accession: AB6036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <STO>
A:Cross-references: GB:AC005174; NID:G12518349; PIND:AKG58749.1; GSPDB:GN00145; UWGP:Z5029
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z5029

```

Query Match	9.2%;	Score 986;	DB 2;	length 1588;
Best Local Similarity	22.9%;	Pred. No. 1.3e-30;		
Matches 510;	Conservative 273;	Mismatches 696;	Indels 752;	Gaps 95;

[illegible]

Dd	508	T-----	NLDTSVGDIG-ADALLMNEKKAFSA--	534
Oy	636	FAGDGPWYIKKEJELLKVKGKTADIDLTKNNIGVADSDNSLTVKLATLSDDLAVN	755	
Dd	535	--HGDDTSEK--LTNVKDA-----DLT-----	ADST-AANGSOLTTTNDAVTN 574	
Oy	736	TKTLASDKYVDSGNNTALONGDLTFESQNTGAPALNSTIGVDGLKFTDNNGI--A	813	
Dd	575	TTNI-----ANNTSMIAINTNTNISLAETV----	NLEDALKMDKDGVFTA 618	
Oy	814	IDGTYYITRKDFVGAKODGSIDLK-SKPYLDKRLKGEVEITTINGINGKAITCISLTL	872	
Dd	619	AHGLE--TTSKITVWK-DGDLTGTSIDVAWSOLKTTNNAVATNTNTNATNT-TNISULT	674	
Oy	873	JTDATN-----ATTGHVTOLGIVSDTKTRAASIGDVINAG--FNLKNNG	914	
Dd	675	EIVYNILGEDALKMKDKONGVFTHAAGNNTASKITINIDGTVFATSSDAINGSLDYLSN-	733	
Oy	915	DAKDEVSFYDVIDFINGNMTAKVYTIDGKASKAYIVYNDGTTIHILTGAONKNOIGYKT	974	
Dd	734	----IATY-----FGGNAS-----VNDDGVFTGPYKIGETINYV--	765	
Oy	975	TLTRKTBKAGD-KAIN-FSVNSGDKALIANKDIADLNLTAGEIRNTGTADTALOT	1030	
Dd	766	-----GALAIAINSFSTSIGD--ALL-----WDATACKFSAKHST-----	799	
Oy	1031	FQVKKVKEGNC-----DDNDADTTYGDKAKTNOVUTLKLGN-----GLDIQTNRD	1078	
Dd	800	-----NGSASITYVADGEI-----SOSSDAVNGSOLDHVSSSYVDALGGAEVNAD	847	
Oy	1079	GTVTFGINTOSGLKAGNNTLLNNG--LSIKNTAGEOIQVGADGYKFAKANNYGAGCI	1136	
Dd	848	GTTTPPYTYIA-----NADYDVWGVALMAIDTTLDDALIMPAD-----	885	
Oy	1137	DGTRIFRDELFEGACTNGSIDSKSPHLKSGCIAAGSKITINIOSEIAONSNDAYTGKI	1196	
Dd	886	-----AGEGAR--SAAR-----GRKRTASVTITNANGAISASSADAINGSOL	926	
Oy	1197	YDLKTELENKISSAKRAON---SLHEFSVADOGCNFPVSNPSSYSPTSIDVTFPAG	1253	
Dd	927	YTTNKYIADALGGAEVNAGGITAPITYITIANEYNN-----VGDALDMD	972	
Oy	1234	ENGITTYKVNKGVRBGIDQTKGLTTPKLYIGNNGKGIYIDSONGONTITGLSNTLANVT	1313	
Dd	973	DNALLM-----DETA-----NGAGAVNASHDCKASI-----ITNVA	1004	
Oy	1314	NDKSGVRTBEGKLIKREDBTKRASIVDYLSAGFINOGENEAVDPEVSYDTYNFADGNAT	1373	
Dd	1005	N-GSI-----SEDST-----DAVNGSQLMAT	1024	
Oy	1374	TAKVYIDDTSKTSKVYDY--NYDDTIE-----VKDKILGYVTTTTSTGANKF	1423	
Dd	1025	NNMT-----EONTIOINOAGNTDATTIIONGACIYVFRINDDGFLAFNDASQSIGATAI	1079	
Oy	1424	ALSNOATGDALV-----KASDIYAHLNTLSGDIOT--AKGASQANSAGYVADGNKYI	1475	
Dd	1080	GYNVSAKXDSVAIGQSYPVDIGIALIGSSVSSRYIAKKSBDNS-----ITEGVVIG	1134	
Oy	1476	YDSDTNKRYQAUKNDGYDKTEVAKDKLVAAQOTPDGTLAOMNVKSVINKEDVANDANKKO	1535	
Dd	1133	YDTTGELGALSIG-----DDGKYRO-----	1156	
Oy	1536	GINEDNAFVKGLEKXAASNKRKMAAVVGDL-----NAVATPLTF--AGDGTTFARKIGE	1589	
Dd	1157	IIN-----VADGSEAHDAVYIROLONAICAVATTPPKYFIHANSTEDSLAVGT	1204	
Oy	1590	TLTTCGCTDPNKLTDNNIGNVAGT-----DGFYVLAKDLITNLNSVNAAGTIKIDKG	1642	
Dd	1205	DSLAMGATIIYN--GDGIGIGYCAVYDANAALGIAIGSMAOYIHNVSIAIGNSTTTRG	1262	
Oy	1643	VSF-----VSSQOKANTPIVYLSANGLDIGGVYISNGVKGRKDPDANAOULINEVRML	1696	
Dd	1263	AQTNTATNMOPANSVGEFSVGSD-----GQRQIITNVAGSADDYDANVVGL-----	1311	

Oy 1697 GUGNAGNDNADONOVNIADIKKDPNSGSSNRRTYKAGTVLGKGNNDTEKLTAGVOVG 1756  
 Db 1312 -----KVDAQVSQNTOSTIN-----LDNRVTNIDSRVT--NIENG 1345  
 Oy 1757 VDKDNANGDLSNVMVTKQKDGSKKALLATYNAAGOTNYLTNNPAEALDRINEOIGREFH 1816  
 Db 1346 I-----GDIY-----TTGSTKYFKTN----- 1361  
 Oy 1817 VNDGNOEYVQGRNGIDSSAGSKHVAIGFOAKADGEAAVAIGRTQGNOSIAIGNAQ 1876  
 Db 1362 -----TDGVDASAGKDSVAL-----GSGSIAADN-- 1387  
 Oy 1877 ATGDOSIAGTGNVYAGKHSAGIADPSTVKADNSYVGNNOFTDQTDFGVGNNTIV 1936  
 Db 1388 -----SVALGTSV-----ATEENITISVGSSTNQ 1411  
 Oy 1937 TESNSVALGSNSAISAGTHAGTQAK-KSDGTAGTATTAGATGYKGFAGQTAAGVAVSYGA 1995  
 Db 1412 RRTTVAAAGKNA-----TDAAVAVAOLKSEAGVRYDTRKADGSDID--YSNITLG---GG 1460  
 Oy 1996 SGAERIONVAAGEVSATSTDAVNGSOLYKATOGIANNTN---ELDRHIONENKANAG 2051  
 Db 1461 NGGTRISVWSAG---VNNNDVYNTAOLKOSYQETKQYTDORAVEMDKLSTESKLSGG 1517  
 Oy 2052 ISSAMAMAMPQATIPGRSMVTGCIATHNGOGAVALSKLSDNGQWFEKINGSADTQGH 2111  
 Db 1518 IASAMAMGLPQATYTPGASMSIGGTYNGESAVALGVSMVANSRMYWTKLOGSTINSQGE 1577  
 Oy 2112 VCAAVAGCFHF 2122  
 Db 1578 YSAAAGAGIOW 1588

RESULT 3  
 H91188  
 probable adhesin ECS4480 [similarity] - Escherichia coli (strain O157:H7, substrain RIMC  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 02-Nov-2001  
 C:Accession: H91188  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yochoyama, K.; Han, C.G.;  
 gisawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A:Reference number: A9629; MUID:21156231; PMID:1158796  
 A:Accession: H91188  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1588 <NAV>  
 A:Cross-References: GB:BA000007, PIDN:BA837903.1; PID:91363955; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECS4480

Query Match 9.2%; Score 986; DB 2; Length 1588;  
 Best Local Similarity 22.9%; Pred. No. 1.3e-30;  
 Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

Oy 1 MHHTKVTNKATGTFMAVAEYAKSHSTGGSCATGQVGSVTLSPARIALAVYIGAT 60  
 Db 1 MKRIKFIWNPATGNTVTSFETAKSRGKRSK-----LILSALVAGM 45  
 Oy 61 LNSGAVAGISSEADGKG-----GANRQDKSIAIGDIAOALGOSIASIADNKTIVNSNN 117  
 Db 46 L--SSGALANAGNDGCGVYDYGSGSAGDGVAIGKAKA-----MFMNTSSSS 93  
 Oy 118 ANIG--AKASGNESIAIGDVLASGHASIAIGSDLLYKKTETVOISOELPLIRGQKALN 175  
 Db 94 TAVGDAIAIEGQYSSAIGSKTHAIGASWAF-----VSAISE----- 131  
 Oy 176 DTYQLADTLMQYRRTHAGHASTAVGAMSYAKGHFSNAFGRTAIAEGYISLAVGLTATA 235

Db 132 -----GDRSIALGASSYSLGQYSMALGRYSKALGLSLAMGSSKA 172  
 Oy 236 KAASIVAGSNAOAI GFATAVAGSTOVNLRNGIALGFGSOVLQKNDPNAANVAYAPD 295  
 Db 173 EGANAIALGNATKATELMSIALDGTANASKAYSMALG-ASSVASEENAI-ALGAELEAE 230  
 Oy 296 DNOPIIDNRKATFRNG-----ATDVESIGNSGNDISIRRIINVAGASADTAHVNA 347  
 Db 231 NATAIGNNAKAKGNSWAMGFGSLADKVNITIALGNSQALADNAIAIGOG-----NKA 283  
 Oy 348 QLEAVVLANRQITFGDSDNNRVEKGLKTLITGGAQTSALTIDNIGYONGDLKAYQ 407  
 Db 284 DGYDALALGN-----GSOS-----KELN--TIALGTSNATGDKSLAGSNS----- 324  
 Oy 408 LAETILSKAVTTEENILANKVTVGKTRITTDKIGFTNDMGIDESKPYLDKDTG--IHA 465  
 Db 325 -ANGNSVALGADSIADLDVTYVGNSSL--KRKIYVKNKGAIKSDSY-DAINGSQLVA 379  
 Oy 466 GQOKITKTAG--VDDDAATYQOLKKNQTAESALQFTVKKVKNKNDAN---DSKI 519  
 Db 380 ISDSVAKRLGGAAVDVDDGT-----TAPYTLNKGSKNNYGALAVIDENT 427  
 Oy 520 I---TYGKNNKBDGQVNTLKLKGENGVDYTTETGTVTFGLNQNNGLTGNSLTNDG 575  
 Db 428 LQWDQTKGYSAAHGTSPTASV-----TIDVADGTLIS----- 460  
 Oy 576 LSVKNTNSNQIOYGADGITEFTDISNSKPGAGIENTRITRDIGRANNTGLDANKPRL 635  
 Db 461 ASSKDAVNGSOLKATNDVE-----ANANATNTSNATNTANATATWTNTI 507  
 Oy 636 TPTGINAGKELINVOGSAIPATNGOLDPMNRSLSTANTEKSGSAATIKDLYLSQVPLT 695  
 Db 508 T-----NLTVSVGLQ--ADALLNNEKKAFAA----- 534  
 Oy 696 FAGDTGPNVTKKLGEILKYKGGKTTADLTKNNIGVADSTDSILYKLAITLSDAVN 755  
 Db 535 ---HGDDTTSK--LTNVSDA-----DLT-----ADSTD--AVNGSOLKTTNDVAIIN 574  
 Oy 756 TKTLTASDKVTVDSGNNTAKLQNGDLTFSSKONTGAPATNSKITGVGDKFTDNNGI--A 813  
 Db 575 TTNI-----ANNTSNLATNTNTNISNLTEVT-----NLGBDALKWMDKNGVEFTA 618  
 Oy 814 LDGTTTYITKQKVGAKDGSIDK-SKPYLDKMLKGEVITTINGINAGKALTIGSLNLT 872  
 Db 619 AHGTE--TTSKITNVK--DGLTLTGSDVAVNGSOLKTTNDVAANTNTINATNT--TNSNLT 674  
 Oy 873 TDATN-----ATTGHVTOGIYDSTPDKTRASIGDIYINAG--FNLKNG 914  
 Db 675 ETVTNLGEDALKWMDKNGVFTAAHGNNTASKITINILDTYATSSDAINSOLYDLSN- 733  
 Oy 915 DAKDFVSTYDVTVDINGNATTAKYVDGKASKVAYDVNDGTTIHLTGADGNKNOIGVKT 974  
 Db 734 ---IATY---FGGNAS-----VNTDGVFTGPPYKIGETNNYV-- 765  
 Oy 975 TTLTKTAKGD--KAIN--TSVNSGDDKALINKDIADNLTLAGELRNKGTADALQT 1030  
 Db 766 -----GBALAAINSSFTSLGD--ALL-----WDATRAKFSAKHGT----- 799  
 Oy 1031 FOYKVKENGQ-----DNDADRTTVGDAKTQVNTLTKLKGK-----GLDIOFNKD 1078  
 Db 800 ---NDGASVITTVADGEI-----SDSSDAVNSQLHGVSSVYVDLGGGAEVNMD 847  
 Oy 1079 GTVTEGINTOSGLKAGNNTLNNNG--LSIKNTAGNQIOVGADGVFAKVNNGVYAGI 1136  
 Db 848 GTTAPATYTTA-----NADYDNGDALNIDTLDALAMDAD----- 885  
 Oy 1137 DGTTRITRDELGEFGCTGSLDKSPHLSKQGINAGKKTINIOSGELAONSNDAYVGKI 1196  
 Db 886 -----AGENGAF--SAAH-----GKDKTASVITINVANGAISASSDAINSOL 926  
 Oy 1157 YDKTELENKISSYAKTAON---SLHEFSVADEGNGFTVSNPYSSYDTSKTDIVTFPAG 1253  
 Db 927 YTTKKYIADALGGAEVNADGTTAPYTTIANAEYN-----YDADALDAD 972

QY 1254 ENGTTTKVKKGVVGVGIDOTKGLTTPKLTGVGNNNGKGIYDSONGONTITGLSNTLANVT 1313  
 Db 973 DALLM-----DELTA-----NGSAGAVMASHDKASI-----ITVNA 1004  
 QY 1314 NKGSGVRTTEOGKTIKDEKTRASIVDLVSLAGFNLQNGEAVDVTSTYDVNFADGNAT 1373  
 Db 1005 N--GSI-----SEBST-----DAVNSQNLNAT 1024  
 QY 1374 TAAVYTDTSKTSKVVYDV--NVDDFTLE-----VKDKLVKTTTLTSTGANKF 1423  
 Db 1025 NMMI-----EQMTQIINOLAGNTDAFYIOENGAGINVTNTDGLAFNDASAGVGATAI 1079  
 QY 1424 ALSNATGALV-----KASDLVAHLNLTSLDIQ-----AKGSQANSSAGYDADGNVI 1475  
 Db 1080 GYNVSAKGDSSVAIGOGSYSDVDTGIALGSSSVSRVIAKGRDTS-----ITENGVIIG 1134  
 QY 1476 YDSTDKVYQAKNDGTVDKTEKAVAKDLVAQAQTPDGLAQMNVKSVINKEQVNDANKQ 1535  
 Db 1135 YDTTDELLGALSIG-----DDGKTRQ 1156  
 QY 1536 GINEDNAFVYKLEKASDNKTRNAAYVGD-----NAVAQPLTF--AGDTGTAKKIGE 1589  
 Db 1157 IIN-----VADSEAHDAVTVROLNAILGAVATPTPKYFHANSTEDSLAVGT 1204  
 QY 1590 TLTIKGGDTNKLTDNNNGVAGT-----DGFYKLANKLJNLNSNAGTKRIDKG 1642  
 Db 1205 DSLAMGAKTIV--GDKGIGIGAYVDANALNGIALIGSNAOVHVNSTAIKNGSTTTKG 1262  
 QY 1643 VSF-----VDSGQAKANTPVLSANGDLGKGVISNNGKGTDPDPAANVOOLNEVRNL 1696  
 Db 1263 AQTNTAYAMMDAPONSVEGFSVGSAD--GGRQITNVAASDDDAVAVGQL----- 1311  
 QY 1697 GLGNAGNDNAGNQVNIADIKKDPNSGSSNRTVYKAGTVLGKGNNDTEKATGCVQVG 1756  
 Db 1312 -----KVTDAQVSOMTOSITN-----LDNRVYNLDSRYT--NIENG 1345  
 QY 1757 VDKDGNANDLSNVWVKTKQDKSKKALLATYNAAGOTNYLTNNPAALDRINQGRFPH 1816  
 Db 1346 I-----GDIY-----TTGSKFKFKTN----- 1361  
 QY 1817 VNDNOEPVVOGRNGIDSSASGKHSVAIGFOAKADEAAVVAIGRQTOAGNOSIATGDNQ 1876  
 Db 1362 -----TDVDAASAGQKDSVAI-----GSGSIADAN-- 1387  
 QY 1877 ATGDGSIATGTGNVAGKHSAGIADPSTYKADNSYSVGNNOFTDATQTDVFGVGNNTV 1936  
 Db 1388 -----SVALCTGSV-----ATEENTISVGSSTNQ 1411  
 QY 1937 TESNSVALGSNSAISAGTHAGTQAK--KSDGTAGTTTATAGTGVKGFAGOTAVGANSVGA 1995  
 Db 1412 RRTTNVAAGKNA-----TDAVNVVAQLKSSBAGGVRYDTRKDGSLD--YENITLG--GG 1460  
 QY 1996 SGAERRIONVAAAGEVASTSDAVNGSOLYKATGIANATN---ELDRHIHONENKANAG 2051  
 Db 1461 NGGTTTISNVSAG--VNNNDVYVNAQLOKOSVQETKQYTDQRMVENDKNLKSKESLSGG 1517  
 QY 2052 ISSAMAMASPPQAIIPERSKVTGIGIATHNGGAVANGSLKSLDNGQWVKINGSADTQCH 2111  
 Db 1518 ISAMAMATGLPQAYTPGASMASIGGTYNGESAVALGVSVANSANGRMVYKLOGSTNSOGE 1577  
 QY 2112 VGAAVGAGFHF 2122  
 Db 1578 YSAAALGAGIOW 1588

RESULT 4  
 A82615  
 surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: A82615  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82615; MIMD:20365717  
 A:Note: For a complete list of authors see reference number A59328 below  
 A:Accession: A82615  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1190 <Sim>  
 A:Cross-references: GB:AE004017; GB:AE003849; MID:99107083; PIDN:AAF84783.1; GSPDB:GN  
 A:Experimental source: strain 9a5c  
 R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre  
 as-Vello, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.  
 Submitted to Genbank, June 2000  
 A:Authors: Ferreira, V.C.A.; Perito, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr  
 J.D.; Jungelira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri,  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
 M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1981

Query Match 6.88; Score 728; DB 2; Length 1190;  
 Best local similarity 23.33; Pred. No. 8.5e-21;  
 Matches 354; Conservative 191; Mismatches 516; Indels 458; Gaps 59;

QY 707 KLGELTKYKGGKTTADDLTKNNIGVVDSTNSLTVKLAKTISLDLAVNFKTLFASDPVT 766  
 Db 27 KIGGSLQ--HAWMTSAASSK--GTQPRRNNAKTAHROHRLHLLVILVLAAT 82  
 QY 767 VDSGNNFAKLO--NGDLT-----FSKONTGAPATNSKT-----IGVDLGFPTDNNGIA 813  
 Db 63 GYTGKVAQVYVNSDSTENCVEILGSSQTSFHSASNDKCKPDTQGEYSLFYDRNLV 142  
 QY 814 LDGTYITRKVKYKQDSLDKSKPYLDKRLKVEYEITNNGKRAITGLSNTLT 873  
 Db 143 LGGSLYVNEGKLGIDISGA-----TYSMRGSI-ATMNG--SAGIDSIAGSOGS 191  
 QY 874 DATNATGTHVQLGIVSDTDKTRASTG-----DVLNAGFNKLNKGDAKDFVSTYDT 925  
 Db 192 KTDGNTSGATVAAQGL-----RSTAICTIARSOQDAISIGTGASTGTAIAIGCAL 244  
 QY 926 VDFINGNATTAATKATYDGRASKAVAYDVNDGTTILHTGADGNKKNQIGVYTTTLFTKDAKD 985  
 Db 245 TSIANGIALGASSVTTTGGVALAGGGLAATASGITVD-----PVTKSTS-- 290  
 QY 986 KAINFSVNSGDDKALINAKDIADINLNTLAGEIRNTKTA---DT--ALCTQVKKVKEG 1040  
 Db 291 -----TLTSMWRSTLGAVSIG--NITSSSTQTRQTLTGLAASRSDPAVNAQLKL--- 340  
 QY 1041 DDDNDADITTVGKDAKTQNOVNTLKLKGNGLDIOTKNGKQYTFGINTSGKAGN--NTTL 1099  
 Db 341 -----AESGGGWNLTASGANSNVALEBSYDLK--NSGNLITTTTSDNVTPLAAL 394  
 QY 1100 NNNGLSINKTAGNEQIOVGADVGRFPAKVNNGVYVAGIDGTTTRTFDELIGFAGTNGSLDKS 1159  
 Db 395 KYDSLTLTGNTA-----MTDGTGVGK-----RVLDSTGLVIAEG----- 429  
 QY 1160 KPHLSKDGINAGGKITTINQSEIINQNSNDVATGKTIIDKTELENKISSAKTAQNSLH 1219  
 Db 430 -PSVVISGINAGQKIMNVTGTA---DTDAVNFQGL-----QAVSDTA----- 469  
 QY 1220 EFSVADGEGNNFTVNPSSVYDTSKTSVDVITPAGENGITTKVNGGVVAGIDQTKGLTTP 1279  
 Db 470 ----- 469  
 QY 1280 KLTGVNNGKGIYDSONGONTITGLSNTLANVTNDKGSVTTGEGTIKDEKTRASI 1339

```

Db 470 -----SKGMILLASGRTSSNVAPGASVDLKN-----TQGNLLI-----TKIAGI 508
Qy 1340 VDVLASAGNLOGNEADVFSVTYTVNFADGNATTAKYTTIDDISKTSKVYDVVDDTTI 1399
Db 509 NDVT---FNLATALEA-DSLTTGNTAMTDDVTGVGSVNT---LQSTGVLITD----- 553
Qy 1400 EVKDKKLGVKTTTLTSTGTGANKFALSNOATGDA---LYKASDIYVHLNLTLS-GDIOJAK 1455
Db 554 -----GFSYTSISGAGNOKITNVAGTADTDAVNFSQLAVSSTASGKWNLLAS 603
Qy 1456 GASQAN---SSAGYVADAGNKVITYDSTDNKYYQAKNDGTVDKTEKVAOKLVA--CAQI 1509
Db 604 GANSNNVVPGESVDLKNDSGNLLITKTTDS-----NDVFENLATALKVDLSLTGNTAMT 657
Qy 1510 PDGTLAONNYSVINKEDVNDANK--KQGINEDNAFVKGLEKASDNKTKNAATVVDLN 1567
Db 658 TDGVTVGSNVTLLGSTGLVITDGPSTSSGI-----SAGNOKITNVAGTADTD 705
Qy 1568 AVAOTPLTFAGDTGTAKKL-----GELLTIKGOTDTNKLTDNNIGV--VAGTDGFTVKL 1621
Db 706 AVNFSQLOAVSSTASGKWNLLASGANSNVAPGESVDLKNITVTSKESGSDVLEPNL 765
Qy 1622 AKDITNINSVAGGTKIDDKGVSE---VDSGQAKANTPVLSANGLDGGKVISNVGK 1676
Db 766 SSSL-KIDKLTIVGDTVMTNGVTVGSGVTLGSMGLVITDGPSTSSGINAGSKITNVAA 824
Qy 1677 GTKRTDANAOQLNEVFNLLGLGAGNADNAGNOVNIADIKKDPNSGSSNRTVYKAGTV 1736
Db 825 GTADTDVNNLSL-----TA 840
Qy 1737 LGGKGNNDTEKLTAT--GGVGVGVVDKNGANDLSNVWVKTOKOSKALLATYNAAGQTN 1794
Db 841 MAGSAGSVHYSYIDGDTQ-----GQNYND 867
Qy 1795 YLINNPALADRINEGIRFPHVNDQNEPVYOGNGIDSSAGKHSVAIGFOAKADGEA 1854
Db 868 -----GATGTRSIAGVGLASAG 887
Qy 1855 AVAIGRQTOA--GNQIAGIDNAQATGDSIAGTG-----NVVAGKHSG----- 1897
Db 888 ATTAGSAAASGKSTIAGNAVASADGVALDGGAKDARGAESYTGKYSGLQNNVTGT 947
Qy 1898 -AIGDPSTVADNSYSVGNNOFTDAT---QTDVFG-----VGNNI-----TYESN 1940
Db 948 VSVGDASKGFRVSNVADAKKEATDAVNLQDLRVADANRYDNKTESSEBQTEYKVN 1007
Qy 1941 SVALGSNSA--ISAGTHA-----GTQAKSDGTA-GTTTAGATGYVKGFGAGTAVG-- 1989
Db 1008 SL---NNSAPPIAGADVATAIGVATASGADSIAMGKASASADNAV-----AIGNHS 1057
Qy 1990 -----AVSYGASAEERRIGNVAEVSATSTDAVNSQYKXTGOGAANNTLDRIRH 2043
Db 1058 VADRANTVSVGSAGSEKQVTVNVAAG---TADTDAVNSQL---NOGLITAKQYTTDVS 1111
Qy 2044 NENKANAGISSAMAMAMPQATVIGRSMTGTGATNNGOGAVAVGLSLDNGQWFEKIN 2103
Db 1112 LRRDTGCGVAAALATLANLPAYIIPGGMISVGSVYRGOSALIVGVSSVSESQWYFKFS 1171
Qy 2104 GSADTQCHGCAVAGCFHF 2122
Db 1172 GSANTRSQVGIGAGVGYOW 1190

```

## RESULT 5

AC0976  
 Probable autotransporter sapB [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AC0976  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churruarin, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

```

S; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1107 (P&R)
A:Cross-References: GB:AL13382; PIDN:CAD03303.1; PID:G16504923; GSPDB:GN00176
C:Genetics:
A:Gene: sapB

Query Match 6.3%; Score 670; DB 2; Length 1107;
Best Local Similarity 23.7%; Pred. No. 1.4e-18;
Matches 1304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;

Qy 993 NSGDDKALLT-NAKDIADN---LNTLAGEIRMTKGA---DTALDTFQYKKKENGDDDN 1044
Db 46 NAGNDGSIQNTTDAITNTTISNNLSNVTLTDLALMDAASGTFSSAR--NGSASK 102
Qy 1045 DADTI--TVGKDAKTQOVN-----TLKLGKNGLDIQTNKDGTVPFGINTQSGKAGNN 1096
Db 103 ITNLAAGTLAAS-IDAAGSQLFETNEKVDQNTADITN-----TNSIQNTTDLATNT 156
Qy 1097 TITLNNGLSIKNTAGNEQIQVAGADGVKFAKVVNGVYAGAGIDGTIRITNDEIGFAGTNGSL 1156
Db 157 TSINN-----LSNSV-----TLTDLALMDAASGTF 183
Qy 1157 DSKPHLSKDGINAGGKRTITNIOGFIQNSNDAYTGKTYDLEKLENNKISSTKATQON 1216
Db 184 SASR-----NGSASKITNLAAGTLAASIDAAGSQLFETNEKVDQNTADITN-----TADITNTN 234
Qy 1217 SLHEFSVADQGNNTFVSNPYSSYDPTSKTSVYI-----TF-AGEGNTTKV--NKGVYR 1267
Db 235 SINQNT--DIATNTTSINLSNVTLTDLALMDADSGTFSSARNSASKITNLAAGT 292
Qy 1268 VGIQDTKGLITTKLVGNNNGKGIYIDSON--GQNT--ITGSLNTLANVTNDKGSVRTTEQ 1324
Db 293 LAADSTDAV-----NGSOLYETNOKYDQNTSAIADINTSTNLSQD----- 333
Qy 1325 GKIKEDKTRAAIYDVLSAGFNLOGNEADVFSVTYTVNFADGNATTAKYTYDDTSK 1384
Db 334 -----NLSMN-----ETTNSASAGSSTTKITNVAAAG 363
Qy 1385 TSKVYDVVNDPTIEVKDKLGKVTTLTSTGTGANKFALSNOATGDAVYKASDIYVHL 1444
Db 364 LSEESTDA--VNSQLFETNEKVDQNTDIAANTN-----ITQNSTAIENL 408
Qy 1445 NTLSDIQAKASQANSAGVVDAGNKVITYDSTDNKYYQAKNDGTVDKTEKVAOKLKV 1504
Db 409 NTSVSDINT-----SITGLTD---NALMDE--DTGAFSANHGGSITKTNVA----- 451
Qy 1505 AQAOTPDGTLAONNYSVINKEDVNDANKKQGINEDNAFVKGLEKASDNKTKNAATVYV 1564
Db 452 AALSDST-----DAVNSQLYETNOK-----VDQNTS-----AIA 483
Qy 1565 DINAAVQPTPLTFAGDTGTAKKLGETLTIKGGQDPTNKLTDNNIGVAGTGTGTVKLAKD 1624
Db 484 DIN---TSITNLTGTDALSMWDEGAFSAGTSGTNTITN-----VAAQ-----ELASD 529
Qy 1625 LTNLSVNAAGTKIDDKGV---SFYDSSGQAKANT--PVLSANGLDGGKVISNVGKGT 1679
Db 530 ST--DAIN--GSQLEYETMLISQYNESISQLAGDTSETIYIENG--TGKVIIRINDGLE 583
Qy 1680 DTD-----ANVOOLNEVRNLLGNGNNDNADGNOVNIADIKKDPNSGSSNRTVYK 1733
Db 584 GQDAVATNGATVAGYDAVAGAGLALGQSSSSIESBIA-----LGSSTSRRAITTG 638
Qy 1734 -----GTVLGKGNNDTEKLTATGCGVGVVDKNGANDLSNVWVKTOKOSKALL 1784
Db 639 IRETATSQGVYI---GYNTDRELLGALSLGT--DGESYQIINV-----ADGSE----- 684

```

QY 1785 ATYNAAGTNTLNNPAEALDRINEGIRFPHVNDGNOEPYVGRNGIDSSAGKHSVAI 1844  
 Db 685 -----ADAVTAVTOLNVAIGAVTTPTKKYHANSTEE-----DSLAVGTSLAM 728  
 QY 1845 GFOAKADEAVALIGRQTA-----GNOSIAIDNAQATGDSOIAIGTVNVAGKHSAGIAD 1901  
 Db 729 GAKTIVNADAGIGIGLNTLVNADALINGIAISNARAHANSIANGNSQTT---RGQOTD 785  
 QY 1902 PSTYKAD-----NSYVGNNN---QFTD-----ATQDVEGVC-----NNITVT 1937  
 Db 786 YTVANMDTPQNSVGEFVSSEDGQRLTNVAAGSADPDVAVGOLKTVDAQVSRNTQSIT 845  
 QY 1938 ESN-----SVAGSASISACT--HAGTQAKSDGTA-GTTTTACATGTWGA 1993  
 Db 846 NLNTOVSNLDRVTNIENGIDIVTGTSTKFKTNTDGDADNAGCADSVAIGSGIAAE 905  
 QY 1984 GOTAVG-----AVSVGASGAERLIONVA-----2007  
 Db 906 NSVALGTSVADDEANTVSVCSTQOQRTTNVAAGVNNTDVAVNAOLKASEAGSVRYETNA 965  
 QY 2008 -----GEVSA--TSTDVANGSQLYATOGIANATN-----ELDR 2040  
 Db 966 DGSVNVSVLNGDGSGETTRIGNVSAVNDPDVAVYAOIKRSEAEANTYTDOKKGEMNSK 1025  
 QY 2041 IHONENKMANAGISSAMAMASPOAYIPGRSMVTGSIATHNOCGAVAVGSLKSLNGQWVF 2100  
 Db 1026 IKIENKMSGGIASAMAAALPQAPAPGAMNTSIAGTFENESAVALIGVSVSGGWY 1085  
 QY 2101 KINGSADTQGHVGAAGVAGFHF 2122  
 Db 1086 KLGTSNSQGDYSAGAGFQW 1107

RESULT 6  
 AD0123  
 probable autotransporter protein yaph [imported] : Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C:Accession: AD0123  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AD0123  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3705 <KUR>  
 A:Cross-references: GB:AL590842; PTDN:CAC89847.1; PID:915979073; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: yaph

Query Match 6.2% Score 669; DB 2: Length 3705;  
 Best Local Similarity 20.6%; Pred. No. 7.3e-18;  
 Matches 593; Conservative 332; Mismatches 1003; Indels 962; Gaps 136;  
 QY 7 VFNKATGTFMAVAEYAKSHSTGGSCATGVGVRTLSFARIALAVLIGATLNGSAY 66  
 Db 365 VIYNGTGSIL-----KATNTGILATKNNANMSDI-YIRSA-----GDTT 403  
 QY 67 AGIGISEADGCGGANAKGDKSIAIGDIAQALGSOSIA-----TGDKNIVHNSNNANIG 121  
 Db 404 AATGISATGHTGTVKIKNDGTTTGTAGIAISSAKIEISVDNDDGTTTATAGTGVNL 463  
 QY 122 ARA-----SGNESIAIG-----DVLASGASIASIGSDDLXL 153  
 Db 464 ASAILNLFEGTINTSKATNGITFAGTE-----GGHLLTDLTTLIGTGLALSNAVAVNLT 519  
 QY 154 KREYVOISLPLIRGKALNDIYQADTNLQKRYRTHAOGHASTAVGAMSYAKGHFSN 213

Db 520 SNVTNLTN-----GALNLSTGLTLVDSLNGHTINIEBAGIGIAATNTE-----LN 567  
 QY 214 AFGTRATBEETSLAVGLRATA-----KAASSIANGNAAIGRA-----ATVAGGS 260  
 Db 568 TFDPAEALDIVNAGAGIGIQTGGVNLASNLII-VVANTGLTALQITDGIQNTTITIGNE 626  
 QY 261 TOYLNLRGIA--LGFSGOVLQKQNDVNANAVRAPD-----295  
 Db 627 IQLAENATRIINFLGSSSKTLNNNGTIIKSVIPAGVADHIINNGGLDGTITTGAGNDTL 686  
 QY 296 -----DNQPI--DNRYKATEKNGATDVSIGNSGNDISIRKLIINVA--GSAD-- 340  
 Db 687 VLDSSQSNQVNLMDGQNNNSVTIONGAT-VSSIITGNGNDFTIINGMSVSTYGLSIDAG 745  
 QY 341 -----TVAVVAQLKEA-----VRLANROITRKGDSS-----NNREKELGFT 378  
 Db 746 TGLNTLFENASTDELAATSLQGFNTMINLVDSHITTVSDNIGSGWNISSSELLFGST 805  
 QY 379 -----LFTTGAQT--SALTDHNGV-----VONGDLKYOIAETTLSLKM-- 417  
 Db 806 FDGLIHAITLQAGISALVNNNSANVSEQASMFAGTQWONGALTLASNSQGLSAGIIGD 865  
 QY 418 --VTEN-----LFANEKVTVKTRLTDTKIGFTNDANGIDESKPYLDKDTGHIAGQ 468  
 Db 866 GTLNLNDIALFNHVLITNGTLNVAK--NLATTFDFGSTVG-----GAFSGIY 911  
 QY 469 KIRKLTGAVDDDAATYQOLK-KYNQTAESALQTF--TYKVKDKNGN-----DAND 516  
 Db 912 NLKRTTFALADNAALASATLKLSDSVTVTGTDRTLHGLDLSGTLIFDCAVQSQOT 971  
 QY 517 SKITTV-----GKNNKPDGTQVNTLK-----LKENGVDVTE-- 549  
 Db 972 SGVYTVTDLANSCTVNIITSGSMDNDPLATINVTILEQDRAGSTELIINATVGTGIDA 1031  
 QY 550 ---TNGT-VTFGEN-----QNNGLTVGNSTLNDCLSKYNINSKQIOY----- 589  
 Db 1032 LDLLVNGTATITSGTQVQSAIQOQGSTVANA--IHNYGLASSNSGSGLVNVTLSALEL 1090  
 QY 590 ---CAD-----GIFFDIINSKP-GAG-----IENTTIT-- 615  
 Db 1091 LADADALLATESGLTANNRLAELFVGGLVYDAQNGALTLANGSNRYEGTITVYAGE 1150  
 QY 616 -----RDGIGFANNTGSLDANKPRLPTGGINAGKEL 647  
 Db 1151 LILGANGAFQOTSLILDASGASANINGISQTVAGTVTVGIVTIGSGVLTSLGTLNGTIL 1210  
 QY 648 TNVQSAINPAT-----NGQIDFENRLSTANTKSGSAATIKLILYL 669  
 Db 1241 DLITGALNLTGASATVAGGLTGAGTLININGNLS--VSANSGISGOT--HIAVY 1262  
 QY 690 SOYPLTFAGDTPRATYKKIEI-----LKKYGGKT--ADDLTKNN--IGVVA 733  
 Db 1263 ASVTLTDTGTLGTSAVEVLGTLNNGANAMANTVLSGDGIIINRAAVTLGNSNSFGAHQ 1322  
 QY 734 DSTDNSLTVKLAKTLSDLDA-VNTKTLTA-----SDKV-----TVD-- 768  
 Db 1323 ICTGELVELVAGASNLGASAVNNGITLSHLINLVGSESIANVLSGVAAGTVDIIGADT 1382  
 QY 769 --SGNNT-----AKLQNGDLTF--SKONTGATP--ATMSKITGVGKLFPTDNKGALD 815  
 Db 1383 ALTANNSGFLQVYALAGNSKLTVASTNNLNGASSVALAGADTLISLGFNGTGTGNSVTGS 1442  
 QY 816 GTTYITKD-KYFAFKQDG-----SLDKSPYLDKDKLKVGEVETITGNIAGKAITGLS 869  
 Db 1443 GVLQVTTDAEVTLLSSKQVSAVITIDATDLNIDIALNRHVLTGNGL-----LN 1493  
 QY 870 NTLTDATNA-----TTGHTVQLGIVDSTDKTRAASISDVNL-AGFNLKNNGDAKDEPSTY 923  
 Db 1494 VAKNDASIAFPGSTVYGAFS-GIVNLTNTTFALSADNAAALARATKLSDSVTVTGAT 1552  
 QY 924 D-TYDFINGANTYAKVYDGNASK-----VADVAVDGTTHLITGADG--KKNQGVK 973  
 Db 1553 DRLHGLDNLNGT--LTFDGPSPQSQANGVTVYDLDLNLNGSTISITGAGMNEHNPVTPP 1610

```

QY 974 TTTTCTKADG-----KAINFSVSGDDKALINAKDIADMLNLTAGEIRNTKGTADTA 1027
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1611 NVSLLEBOD-RGDILLELINANAVTGNANMLDLVDGTALTSGTQGVESAIQQGGSTVANA 1669
QY 1028 LQTFQKRYKKENDDD-----NDADITTVCKDA--KTNOYNTIKLKGKNG 1071
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1670 IHNYGLTSSNGSGSLVYNYTLALELLANGANMLLTATESGLANRYLMELEFGVGL 1729
QY 1072 DIQTNK-----DGTVT-----FG-----INFOGKAGNN----- 1096
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1730 VYDAQCALTLANGNNRREGTITVTAGELLANGAFQGTSLINTASGASAINNGRYOTV 1789
QY 1097 -----TTLNNGSLIKN-----TAGNEQIQGVADGVKFAKVNNG 1130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1790 GAVTNGAVTLGNGVLTGSLTNGGILDLTGALNLAAGSSVAGLGTAGLTINING 1849
QY 1131 -----VVGAGIDGTTITR-DEIGFAGTNGSLDKRPH-----LSKDDIAGGKKTITNIO 1180
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1850 DLAVSATNGSLGSGQTHIADVASVTLGT-GTLGTSAVEVLGLTINLNGANA--MTNVL 1905
QY 1181 GEIAQNSNDVATGKTYDLKTELENKISSTAKTAQNSLHEFSVADQGN--FTVSNPYS 1238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1906 GGGVITMAAVT-----LSGNNS--FGAHQIGTDELTVGQASN 1943
QY 1239 SYDTSKTSDVITPAGE--NGIT--TKYKGVYRVGIDQTKGLT----- 1278
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1944 LGASSATVNLGLTSLHLINGVSESIANVLGAGSTVDIIGADPALTANNSGFLGQYA 2003
QY 1279 ----PRLYGANNNGK-----IYDSONGNTITGLSNTILANTNDKGSRTTEQKIIID 1330
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2004 LAGNSLTYASTNNLGLASSSVALAGADTLISLGFNGTGNVSGVQLQVYDADAEVTLT 2063
QY 1331 EDKTRAASI-VDVLSAGFNLO-----GNG-----EAVDFVSTY-----DT 1364
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2064 SSNGVGNVYKVDIADATLMLNDIALEFDHVLGTNGTLNVAKNLATAFDEGSGVGAFSOI 2123
QY 1365 VNF-----ADGNAVTAKVY--DDT--SKTSKVYDVNDVDT----- 1398
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2124 VNLNTTFALSDNMAALRAATLKLSDSDVTGVTDRILHGLDLNGTLIFDGSPPQSO 2183
QY 1399 ----IEVKKKILGVTTLTTLSTGTGANK-----FALSNOATGD--ALVYASDVAHL 1444
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2184 ANGCVATVTDALANGSTISTGAGMENEHPVPPNVSLLEQORGLDQLIDADVNTGA 2243
QY 1445 NTL-----SGDIQTAKGASOANSSAGYVADGNKVLYDSTDNKYQAKNDGT---VDKTK 1496
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2244 NDELMINGTTISAGGVSTVQGGGYTANAT-----HNYGTSNGSGGLVYNT- 2294
QY 1497 EVAKDKLVAAQTPDGTLAQMNVKSVYKQVNDANKKOGINEDNAFYKGLKKAASDNKT 1556
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2295 ----LSALELLADGA-----NALLIATESGLTANRE 2321
QY 1557 KNAAYV-VGDLNAAVOT-PLTEA-----GPDGTAKKL-----GET--LIIKGQ 1597
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2322 LNAELSGVGLVYADQNGALTLANGNNRREGTITVTAGELLILGANGAFQGTSLINLASGA 2381
QY 1598 TDTNKLTDNNGVAGTDEFTYKAKDLTNLNS-VNAGGKTIDKGVSVSSGQAK-- 1653
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2382 SANINGRYQTVGAVTMTGVTILNGELSTDTLLINTGMINTD-GILNLENGGASSISG 2440
QY 1654 --ANTPVLSANGIDLGKVIYVNGK--TKDTPAANYQOLNEVRNLLGAGNAGNADNQ 1710
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2441 GLTNGILNIKGGDFTISIDNNGLAGQTNISDGASY-TLNGCTIIGTGNLGS----- 2492
QY 1711 VNIAIDIKDPNSSSSNRTVIVAGTVLGKGNNDPEKLT-----GGVGVGDKN 1762
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2493 -SVIVDLGLNLVADNS-----LANVISGDTIINTIATVTLGSSNSFGAHQIG----- 2540
QY 1763 ANGDSLNVVVKTKQDKSKKALLATYNAAGQTYNL-TNNPAAIDRI----- 1807
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2541 TNGELT-----VGQASNLGASSATVNLGTLTSHLLINGVSESIANVLGASVASTVDIIG 2595

```

```

QY 1808 -----NEGIRFEFHVNDGQEPVQGRNGIDSSAS-----GKHSVAIGF----- 1846
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2596 ADTALTANNSGFLQGYALAGNSKLTVASTNMLGASSSVALLAGTDTLSGPNFCGNSV 2655
QY 1847 -----QAKADGEA-----VAIGQYQAGNSIAIGNAQ 1876
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2656 TGGSVLQVTDIAEYTLTSSNGSVNAVTDIADATLINDIALFHHALTGGLLVAN-- 2713
QY 1877 ATGDOSIAIGTGNVAGKHSAGIDPSTVKADNS-YGVANNQPTDATTQ-----DVFG 1929
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2714 --DASTAFDGRATYGAFTG-----TVNLNNSTFPLDSGNTTVLAQATLKSSGNLTS 2764
QY 1930 VGNNT---VYTESNSVALGSNSAISAG--THACTQAKKSDGTAG----- 1968
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2765 VGNQVONIGTLAMNGTLTLPINIVDNAGILTSDEGTIANNSINTGGEVRYNLPNSLAPS 2824
QY 1969 -----TTTGATGT-----VKFAQTAVGANSVAGS 1997
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2835 LDGLSMELDECEITIVLTATGAATGTGHELTLPDENQDPIASVTVQGVHNGSTSAATG 2884
QY 1998 AERR-----IQNVAGEVSATSTDAV-----NGSOLYK----- 2025
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2885 SPNYGMTTGEDYDGLVYVYGLTLELSTGSEALVTLAILANNQTSNDLSAQITSGDL 2944
QY 2026 -----ATGCIANATNEL-----DRIHQENRKNAGISSAMAMSPQATYIPGRS 2070
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2945 AFASANDSTASLSTNSTSYGTWVSSGNLRDADSA-LGQTSILAMSTATHVDINGTQ 3003
QY 2071 MYTGIAATHNG-----OGAVAVGLSKLSDNQWFKINGSDTOGHV 2112
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3004 QVGELETGEGSTLTDLNDKILTVTGGQIDGALTGGGELVLSG-----LNVSTDNAGFT 3059
QY 2113 GA---AVAGGFH 2121
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3060 GSTDIANCAVAH 3071

RESULT 7
B11704
cell surface antigen (scs3) RP451 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: B11704
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: B11704
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2340 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; MID:g3868717; PIDN:CA14908.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: scs3; RP451

Query Match 6.1%; Score 651.5; DB 2; Length 2340;
Best Local Similarity 21.2%; Pred. No. 1.9e-17;
Matches 519; Conservative 302; Mismatches 879; Indels 743; Gaps 121;

QY 10 NKATGTMAVAE-----YANSHSTG-----GGSCA-----TGQGVSTLSPARIALAVL 55
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 SKIIRKFLATASLGTFTLTSNATGTIIPNNGSVSLNTDAGLVGV-----FNNGDIIQI- 58
QY 56 VIGATNGSAAVAGISIEADGKRGAMARQKSAIDIDNALGOSIAI----- 105
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 -----VWGRIKISADKANAIIGTINTLKE-----LPDFGGEVSEYQVNSIGPLNAGEDLN 109
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 ---GDNKIVHNS-----NNNANIGAK-----ASGNEST 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 TNPRLPFISSNNVSIITGVGKTFNSIDFAGKNATIQINKDLNITTKIDNVTAGNNGSI 169

```



Db 2101 --DNSIVIGAAVTMADSKVHKNDKNGDRTRKAKSNISYIGLYNMLTNFVEALIGVGR 2158  
 QY 1923 -----TQDVFEGVNNITVESNSVALGSSAISACT 1954  
 Db 2159 NKIKNEYKRITITTDQIALGKFINFTYSELGLGYNLISHRT 2201

## RESULT 8

hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1 F90696)

C.Species: Escherichia coli

C.Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C.Accession: F90696

R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A.Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A.Reference number: A99629; MUID:21156231; PMID:11258796

A.Accession: F90696

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-5291 <HAY>

A.Cross-References: GB:BA000007; PIDN:BA33965.1; PID:913360000; GSPDB:GN00154

A.Experimental source: strain O157:H7, substrain R1MD 0509952

C.Genetics:

A.Gene: ECs0542

Query Match 6.0%; Score 645.5; DB 2; Length 5291;  
 Best Local Similarity 21.9%; Pred. No. 9.4e-17;  
 Matches 582; Conservative 293; Mismatches 932; Indels 855; Gaps 131;

QY 35 TCGOVSVRFLSFARIALAVLIGATLN---GSAVAGIGISEADGG-KGANARGAKSI 89  
 Db 1136 SGPPTLTINTVSGDDIINAEIYVAOTISQVGTAVAGNTVYITGGNOYATVGSDSLW 1195  
 QY 90 A1--GDAIALGQSOSIALGDKRYHNSNNANIGAKASGNEISAIIGGDVLAASHASIAI 146  
 Db 1196 SVSVPANVLOALNGELTISAS--LTNSANN---TGTAHVIDIDANLPGRLVDTV 1246  
 QY 147 GSDDLTKKEMVVOISELLPIIRGOKALNDIYOLADTNLQKRRRHAGHASTAY----- 201  
 Db 1247 AGDDVYNSIEHTQAL---VITGSSS-----GLAAGALTVYINSVT 1284  
 QY 202 -GAMSTAKGFHFNACSTRATAE---GTYSLAVGLATATAKASSIANGSAAIGFAATAV 257  
 Db 1285 YGATVLAOSMSGVADVADVTWMPACTVNIASGNTACTTTSI---SHIPVVDLAAYAI 1341  
 QY 258 -----GSGTOYVNLRGIALGFGSQ---VLQKDN---DYNA 286  
 Db 1342 TINTLTDDVINAAEKSDLOLSGTTSGVEAGQITVIFGSKSYTTVAADNTMGLTIPA 1401  
 QY 287 ANVRAYAPDDNDPINRYKATFKNGATDV-FSIGNSNGDSIRKILINVGAS----- 338  
 Db 1402 VDV-----ATLPDGAANYOASVSNVAGSTQATHAYSVDATAFSPVTINT 1445  
 QY 339 -ADTDVAVNAOLKEAVRILNRQ-----ITFKGDDSNRNEVEKGLTITITGG----- 384  
 Db 1446 IATFDILNAAEAGSLTITGTAEGQIVYTLNGVNSGNVQADGMSVSVPFGDLAS 1505  
 QY 385 -----AQISALFDHNIQVYONGDLKVOLETILSKMVTENTL----- 423  
 Db 1506 LPAASVYVNASVDAKARNSASATHN-----LTVDLAAFPVTVINVAAGDITINATEH 1556  
 QY 424 -----TANE-KVTVGKRLTLDKIGFTNOMNGIDESKPYLDKPTGIHAGQ 468  
 Db 1557 GQAQIISGATGATGNTVSVITIGTYT---VLDAAG-----NMSIGVPAS 1601  
 QY 469 KITKTLTAG-----VYDDDAATYQOLKKNVQTAES---ALQFTYKKVKDKNANDSK 518  
 Db 1602 VTSALAGGVITITATYTDGASNGSASHVYVALGAPVLAITITIAVDIINAAEGADLA 1661  
 QY 519 ITTVGKNNKPDGTQVNTLKLKGENGVDTTETNG---TVTFGLNONGL-----TVGNSL 571

Db 1662 I--GTSNQFAGTQI--TVTLNGQ--YTTTADAGSNMVSYPASVSLGERTATVYNAAT 1717  
 QY 572 NNDGLSYAKNINSKQIOVADGDTF---TD-----ISKRGAGIENTRI 614  
 Db 1718 DADGNS--GSASHNVQVNTALPGVTINVAATDDIINAAEGVEQIISQVYGAAGDVTYV 1776  
 QY 615 TRDGIQFANNITGSLDANKPRLTPTGINA---GGKELTNQSAINPATNG----- 661  
 Db 1777 T---LGCATYATVQANISMSVDYPASALOELNGELLISASVTNSGNTGNGTREITID 1833  
 QY 662 -----QIDFNKRSSTANTKESGSAAT---KDLVNLQVPLTFAGDT----- 700  
 Db 1834 ANLGLRVDYVAGDDVYNIIEHGAALYTGSSSLAAGSNVTLTINQTYAVALADGTW 1893  
 QY 701 ---GPNYTK--LGEILIKKVGKTYAD-----DLTKNNT---GVADSTDSL 740  
 Db 1894 SVGPADVDSAMPAGSVYTIASGSGTSAGNPVSYTHPYTVLDSAVAVSINAITADVDVINA 1953  
 QY 741 TVKLAKTLSDLDVANTKTITLASDKVTVDGKNT-----AKLQNGD 780  
 Db 1954 EKGALTLISG---STSGVEAGQTVTVFEGKITYSATVAANGSKVTSVPADMAALRDGD 2009  
 QY 781 LT--FSKQNTGATPATNSKTIQVD-----GLKFTDN----- 810  
 Db 2010 ASAQASVSNVNGSATTTTHAVSVDAAPVTYINTIAGDDILNAAEAGALTTGSSSTAE 2069  
 QY 811 -----GIALDCTTYTKKQVFAKODGSLDKSKPLDKL-----KVG----- 849  
 Db 2070 GQTVYVTLNGTNY---TGTVQDGSWSVSPADSLTASNYTVYNAVSDKAGNPAS 2124  
 QY 850 -----EVEITTTNGINNGKAITGSLNTL---DPTNATGTVTOLGIVDST--- 892  
 Db 2125 VNHILYDTSVPPYTIYTVAGDDVYINATEHQAQIISGATGATGATVYITGINTFTT 2184  
 QY 893 -DKTRAASIG-----DVLNAGFLKNNGDAKDFV-----TYD 924  
 Db 2185 VLDSAGMSVGPASVVSALANGTVTINASTVDAG---GNSGSKHQVYNTGLPTTFEN 2241  
 QY 925 TV---DFINGNATYAKTYDQKASKAVYDVNVOSTTTHLNG-----ADGKNQGVK 973  
 Db 2242 AISGDNILNDEKQPLTISGSGTGLTAQV---YTLNGHNSATTTDASGN-----W 2292  
 QY 974 TTTLTITDKADKAKINFSVNSGGDKALINAKDIADNL-----NTLAGE--IRNT 1020  
 Db 2293 TLTVPVSDLAALGQANTVVSATSAGNTRASGQANLVLVSGLPDVTINTVAGDDIINAA 2352  
 QY 1021 KGTADTALQTFQVKKVKEGDDNDADDTTVGKDAKTNOYNT-----LKIKG 1068  
 Db 2353 EAGADQRTISGVVTRAAA-----GDTVTVLGGNTYATATVQSNLSMSVSVPTADLQALGN 2406  
 QY 1069 NGDLIO---TKKDGTVYFNGINTOSGLKAGNNTLNNN--GLSINKTRAGNQIO----- 1116  
 Db 2407 GDLTITASVTINANG-----NTGSGTR--DITIDALPELRVDYVAGDDIVSIEHQA 2457  
 QY 1117 -----VGADEVKFAKVNNGVAGAGI--DGTTRI-----TRDEIGFA 1150  
 Db 2458 LVITGSSGLNAGAVLVTINSVAYSATVQADGWSGSIIPAANVSAMPAGLVEVDGQS 2517  
 QY 1151 GTNGLSKSKPH-----LSKDCINAGKKTINIOSGELAQ--NSMDAYT--- 1192  
 Db 2518 SANPNVSVSHPTVDTLVAVALISINTVASDDVINAAEKGTMLTISGSGIESQCTVYTF 2577  
 QY 1193 GKKTY-----DLKT-----ELENNISSTAKTAONSLHEFSVADQGN 1229  
 Db 2578 GKKTYTTSVAANGSWSVNPAPADLATTPEGAANYOASVSASGNSASATHAISV--DASAP 2636  
 QY 1230 NFTVSNPYSSYD---TSKTSVITTFAG---ENGITTKVN--KGVVRGIDOTKG---LTT 1278  
 Db 2637 TLTIT--NLIASDILNAAEAGSLTISGTAEGQIVYTLNCAATYTGTVQAGMSVSVS 2695  
 QY 1279 PKLTGNNNGKGIYD---SONGQNTTIGLSNLTANVTNDKGSVRTTEGKTIKDEDKTR 1335



```

Db 2696 PTSLGALNAGNVTYSATVNDKAGNPGSASHNLAVDTTAPVLTINTVAGDDIINDAEHAQ 2755
QY 1336 AASTVDLVSAGFNQNGNEAEDVPS-----TYOTVNPADNONTAKATYTDITKTSKVV 1389
Db 2756 AL-VISGSSG-----GEAGDVVSVVLNGKTYTTTLDASGNWS-----2792
QY 1390 YDVAVDDTTIEVKDKLKKTTTLTSTGTA-----NKFALSNQATGALVKASDIY 1441
Db 2793 -----VVPADVTALGSGAQTITASVSDAGNSDDASRTVYVSLAPV 2836
QY 1442 AHLNLTSGD---IQAKGASQANSAGYVDADGKVIYDSTDNKYQAOKNDGTVDKTEV 1498
Db 2837 ISINTIAGDDVYINATEKSGDLALS--GTSDDQAGTAITVTLNGQVYSAITDASGNWSVTV 2894
QY 1499 AKDKVLAQAQTPDGTLLAQMNVKSVYKQVVDANKKQGINEDNAFVKLEAAADNKTKN 1558
Db 2895 P-----ASAVSALGEATYSVTASVTN-----AQNSTASHNVOVN 2930
QY 1559 AAVTVGDNLAAVAPLTFAGDGTAKKLGELTIKGGQTDNKLTDNNIGVAGTDGFT 1618
Db 2931 TALPBITTPNPAATDDIINASEGSAQTISGOVTAAGASTYVEL-----GKTYT 2961
QY 1619 VKLADLTNLNSVYNGKTRIDKGVSEVSSGQAKANTPVLSANGLDLGRVYSNWKGT 1678
Db 2962 ATVQADLSMNVSVPADMQ-----ALNGELTVNASVYNA-----VGNISGCT 3024
QY 1679 KD--TDAANVQOLNEVRLLGL--GNAGNDAGNOVNIAD-1KKDPSGSSSSKRYIYK 1732
Db 3025 RDTITDA-----SLPGLRVDTVAGD-----VYNIIEHAQAOVITSSSG--FA 3066
QY 1733 AGTVLGGKNDT---EKLATGVGVGVKDGKNGDLSNWKVKQKQKSKALLATATYA 1789
Db 3067 AGTALTIVINNQTATYATVLANGMSVGPAT-----DVSN-W-----PAGTLNITYSGANS 3116
QY 1790 AGQTVYLTN-----NPAEADRIN--EQRIRF-----FHYNDGNEPVVQG 1828
Db 3117 AGTQSTHPLVDLTAVALISNISTSDAINAEKGAALTLSSGSEVAGQVTVTFEG 3176
QY 1829 RKGIDS--SASGRHVV--AIGFOKADGEAAVAIGROTQAGNOSIAIGDAAQATGDOI 1883
Db 3177 GKTYYTTVAANGSMSTVPADLALRDGDASAOV-RVTNNV-----GNSATATHEYSV 3229
QY 1884 -----AIGTNVVAAGKHSGA---IGDPTVKADNSVY---GNNOFFTAOTDYD 1927
Db 3230 DSAAPTYIINTIASDNIINASEAAAGVYSGTSTAQTGTLVTLNGNTYOT--VQTD- 3286
QY 1928 FGVGN-NITVYESNSVALGSNSAISAGT---HAGTOAKKSDG-TAGTT-----TTAG- 1974
Db 3287 ---GMSLTLPRASDLTALANNGYTLTATVSDLAGNLGSAKGVYDTTAPVISEFTVAGD 3343
QY 1975 -----ATGVYKG-----TAGQ-----TAVGAVSYGASGAERRIONV 2005
Db 3344 DVINNVHEIOAIIISGTATGAAGDRLVYTLAGQOYVTLSDASGNWSVGPASV--ISGL 3401
QY 2006 AAGEV--SATSTDAVNGS-----OLYKATQGIANATNEIDHRIHONENKANAGISSAM 2056
Db 3402 AGGTVYISATITDSAGNSSTQTHNOVNTAVALSVSTISGDNLL--NAEAG--SAL 3455
QY 2057 AAMASPMOATIPGRSMVTGCIAT--HNGOGAVALGSLKLSBNGQWFKI-----2102
Db 3456 TLSG-----TGTNFRATGVYTVLLNGKYSAT-----IQSNGSMVNPADVAALSDGT 3505
QY 2103 -----NGSADTQGH 2111
Db 3506 SYTVSASAOADSAGNGNSTOTH 3527

```

```

R:Perna, N.T.; Plunkett, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobleck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MIMD:21074935; PMID:11206551
A:Accession: B85547
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STC>
A:Cross-references: GB:AE005174; NID:g12513366; PIDN:AAAG54838.1; GSPDB:GN00145; UMGF:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 5.9%; Score 634; DB 2; Length 5188;
Best Local Similarity: 22.0%; Pred. No. 2, 6e-16;
Matches 576; Conservative 290; Mismatches 928; Indels 830; Gaps 130;

QY 35 TCGGVSVRTISFARIALAVLVIGATLN-----GSAYAGISISADGG-KGANAARDKSI 89
Db 1136 SGPLTINTVSGDDIINAEIYVAQITISGOVTAAGNTVYITGNGYNATVQDLSW 1195
QY 90 AI--GDIAQALSGSIAIGDNKIVHNSNNNANIGAKASGESIAIGDYLAGSHASIAI 146
Db 1196 SVSVANVLQALNGELTISAS--LTNSANN-----TGTATHDYIDANLPLGRVDYV 1246
QY 147 GSDDLVKKETVQVQISLPIIRGQKALNDIYOLADTNLQKRYTHAOGASTAV-----201
Db 1247 AGDDVVISIEHTQAL-----VITGSSS-----GLAGGALTVVINSVT 1284
QY 202 -GMSYAKGHPSNAFGRATAE--GYSLAVGLTATKAKASSIAGVSNQAIQFATAV 257
Db 1285 YGATVLDGSMVGPVADVTNMPAGVNIJAVSGTNTAGTTISI--SHVYVYDLAAVAI 1341
QY 258 -----GGSTQVNLNRGIALGFSQ-----VLQRDN--DVNA 286
Db 1342 TINTLSTDVYINAEKSDQLSGTSGVEAGQITVIFPGKSVTTVAADNTWGLTIPA 1401
QY 267 ANPRAVAPDDNPIDNKTKATFKKAGADV-FSIGNSGNDISIRKLIIVAGS-----338
Db 1402 VDV-----ATLPDGAANVOASVSNVAGNSTOATHAVSDATPASYTINT 1445
QY 339 -ADDADVNAVQLEKAVYLANRQ-----ITFKGDSNNRVEKLGKTLTIGG--384
Db 1446 IATDDIINAEAGSALITSGTAEACQIVYTLNGVYISGNVOADGSMVSVPTGLAS 1505
QY 385 -----AOTSALTDHNIQVONGDLKQVLAETILSLKMTTENL-----423
Db 1506 LTASSTVYMASVSDKAKNSASATIN-----LTVDLAAPVYTIINTVAGDDIINATEH 1556
QY 424 -----TANE-KYTVGKTRLTJDKIGFTMDNGIDESKPYLDKDTGIIHAGQ 468
Db 1557 GQAQIISGATGATTCMTVSTVITGTYTT-----VIDANG-----NMSIGVPAS 1601
QY 469 KITKLITAG-----VYDDDATYQQLKKVQTAES-----ALQFTYKVKXKNGDANDSK 518
Db 1602 VISALAGDVTITATVDSAGNSGTASHYTVVAGAPVLAINTIAYDDIINAEKADLA 1661
QY 519 IITVGNKRPDGTQVNTLKLKENGVDVTTETNG--TVTFGLNONGNL-----TVGNSTL 571
Db 1662 I--TGTSNQAGTQI--TVTLNGON--YTTTADASGMSVYTPASVSLAGEATYVTAAT 1717
QY 572 NNDGLSVKNTNSKROIQVADGTF-----TD-----ISKRGAGIENETRI 614
Db 1718 PADGNS--GSASHNVQVMTALPGVITNVATDDIINAEAGVEQITISQOVYGAAGDITVY 1776
QY 615 TRDGIQFANNCGSIDANKRPLTPGICNA-----GKELTNVOSAINPFTNGG-----661
Db 1777 F--LGGTTYATVQANLSMSVDYPASALOELGELTISASVYTNVSGNTGNGTRETID 1833
QY 662 -----QIDFNNRLSTANTEKSGSAAT--KDLVNLGVPLTFAGDT-----700

```

RESULT 9  
 B85347  
 Probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: B85547

Db 1834 ANLPLRLVTVAGDDVYVNIIEHQALVITGSSSLGAGSNVTLINGTGYAAVLADGTW 1893  
 QY 701 ---GPNVTKK---LGEILWVKGKKTAD-----DLTKNNI---GVYADSTDNEL 740  
 Db 1894 SVGVPAVDVSAWAGSVTTAAAGSTAGNPVSYTHPEVYDLSAVALAATADVDYINA 1953  
 QY 741 TVKLATLSDLAVALNTKTLTASDKTVDSGNNT-----AKLOND 780  
 Db 1954 EKGALTLTSG---STSGVAGQTVTVFGKTYSAFVAANGSMSTVPAAMALROD 2009  
 QY 781 LT--EKKOTGATPATNSTGTVD-----GLKTTDNN--- 810  
 Db 2010 ASAQASVSVNNGNSATTHAYSDASPTVITNTAGDILNAAEAGALLTIGSSTAE 2069  
 QY 811 ---GIALDGTTYTKDKVGFAPKODSLDKSPYLDKDL-----KVG--- 849  
 Db 2070 GGVYVYTLGNTN---IGYVOTDGSMSVSPSADLSLTLSXYTVNAVSDKAGNPAS 2124  
 QY 850 -----EVEITTINGINAGKATIGLSNTLT---DATNATGHVYQLGIVDST--- 892  
 Db 2125 VHNHLTVDTSVPVYVINTVAGDDVINATEHQAQIISGATGATGATVTVTIGNTFTT 2184  
 QY 893 ---DKTAAASIG-----DVNAGEFNKNDADPVS-----TYD 924  
 Db 2185 VLDASGNMSVGPASVYALANGVTYVINA SVTAG---GNSGSAHQVYVNTGLPFIIFN 2241  
 QY 925 TV---DFINGNATTAKTAVYDGRKASVAYDVNVDGTTIHLTG-----ADGNKNOIGV 973  
 Db 2242 AISGDHILNADEKGOPLTISGSGTGLATGAY---TVLNGHNYSATTDASN-----W 2292  
 QY 974 TTTTAKTAKGKAIKNEFVNSGDDKALINAKDIADNL-----NTLAGE---IRNT 1020  
 Db 2293 TLTVPASDLAALGOANYTVYASATSAAGNTASSQANLWDSGLPDTVINTVAGDDIINA 2352  
 QY 1021 KGTADALOTPOYKKKKEKGGDDNDADTTVGKDAKTNOYNT-----LKLK 1068  
 Db 2353 EKGADDTISGVVTRAA-----GDTVYTLGNTYATVQSNLSMSVSPVPTADLQALGN 2406  
 QY 1069 NGDLID---TNKDGVTVEGINTOSGLKAGNNTLNNN---GLSIRKTAGNEQIO----- 1116  
 Db 2407 GDLITASVTNANG-----NTGSGTR---DITIDANLPLGLAVDVYAGDDYVNSIEHQA 2457  
 QY 1117 -----VGADGVKFAKVNNGVAGI---DGTTRI-----TRDEIGFA 1150  
 Db 2458 LVITGSSGLNAGAVLTVINSVAATVQADGSMVSGIPAAVAMPAGPLTVEVDCQS 2517  
 QY 1151 GNGSLDKSKPH-----LSKDGINAGKKTNTQSGEIQ--NSNDAYT--- 1192  
 Db 2518 SANMPVSVSHPTVDLTAVALISINTVASDDVINAAEKGNTLTSSTGIESGQTVYTF 2577  
 QY 1193 GGRKY-----DLKT-----ELENNKISSTAKTQNSLHEFSVADEQGN 1229  
 Db 2578 GGRKTASVAANGSMVNPADLMTLPBGAAVQASVSGNSASATHAVSV--DANAP 2636  
 QY 1230 NPTVSNPVSSTY---TSKTSVITTAG---ENGITTKVN--KGVVRVGIQDTKG---LWT 1278  
 Db 2637 TLTIT--NTIASDDILNAAEAGSPITISGTASTAGTQVTVTLGATVYTGQADGSMVSYSV 2695  
 QY 1279 PRLTGNNGNGKGIIVD---SONGONTIIGLSTLANVNDKGSVTTDQKIIKEDTR 1335  
 Db 2696 PYSALGALNASVTVSATVNDKAGNPGASHNLAVDTTAPVLTITVAGADDILINAEHQ 2755  
 QY 1336 AASIVDVLSAGFNLOGNEAVDEVS-----TYDTVNFADGNAATKATYDTSKTSKV 1389  
 Db 2756 AL-VISGTS---GEXGDVSVVLNGKTYTTLTASGMS----- 2792  
 QY 1390 YVNVNDOTTIEVKDKKLGKKTTLTSTGTA-----NKFALSNOATGALYKASDIV 1441  
 Db 2793 -----VGVPADVYALGSAQVITASVSDRAGNSDDASRFTVYSLSAPV 2836  
 QY 1442 AHLNTLSGD---IOTAKGSOANSAGVVDAGNKVYIDSTDNKYYOAKNGDVTDKTEV 1498  
 Db 2837 ISINTIAGDVINATEKGDIALS---GTSDDPAGTAITVTLNGQNYSATTTASGMSVTV 2894

QY 1499 AKDKLVAQAPDGLAQMNKSVINKEOVNDAKKOGINEDNAFVGLEKASDNKTN 1558  
 Db 2895 P-----ASAVALGEPATVSYTAVTN-----AOGNSSTASHNOVN 2930  
 QY 1559 AAVTVGDLNAAVQPLPFGDGTGTAKLGETTLIKGQDTEFNKLTNNIGVAGTGT 1618  
 Db 2931 TALPITITINPAFTDILINASEAGSAQITISGQVTAAGSITVEL-----GKTYT 2981  
 QY 1619 VKLAKDLTINLSVNAAGTKIDDKGVSEYDSSGQAKANTPVLSANGDLGKVIYNNKGT 1678  
 Db 2982 ATVQADLSWNVSVPAADMO-----ALNGELTVNASTVNA-----VGNTEGCT 3024  
 QY 1679 KD--TDAAVNOQLNEVRRLGL--GNAQNDNADQONVAD--LKKDPNCGSSSNRTYIK 1732  
 Db 3025 RDTITDA-----SLPELRDYTAGDD-----VYNIIEHAQAOYTIGSSSG---FA 3066  
 QY 1733 AGTVLGGKGNNDT---EKLATGVGVGVKDGANGLDSNVWVTKODGSKKALLATYNA 1789  
 Db 3067 AGTALTVVINNQTAAATVLANGSMVGPAT-----DVSN--W---PAGTLNITYSGANS 3116  
 QY 1790 AGQTYNLTN-----NPAEALDRIN--EGGTRF-----FHVNDGQDEPVYOG 1828  
 Db 3117 AGTQTSIHPPLTVDLTAVALISMNSITSDDAINAAEKAALTLGSGTSGVAGQTVYVTFG 3176  
 QY 1829 RNGIDS---SASGKHSV---AIGFOAKADGEAAVAGHOTOAGNOSIAGNAAOTGDOSI 1883  
 Db 3177 GKTYTYYAANGSMSTVYPAADLALRDGDASAOV--RYTVN-----GNSATRTHEYSV 3229  
 QY 1884 -----AIGGVNVAKRHSGA---IGDPSTVKADNSYSV---GNNQPTDQOTDV 1927  
 Db 3230 DSAAPVYVINTIASNIIINASEAAAGVTSGSTAGQGLTVTLNQTNT--VQTD-- 3286  
 QY 1928 FGVGN--NITVESVVALGMSAISACT---HAGTQAKSDG--TAGT-----TTAG-- 1974  
 Db 3287 ---GSMSTLTPASDLTALANGTYLTFTVSDLAENLGASAKGVTDTTAPVISFNTVAGD 3343  
 QY 1975 -----ATGVKNG-----FAGQ-----TAGAVSVGASGAERIONV 2005  
 Db 3344 DVINNEHIAQIISGRATGAVAGDRLVYVIAQGYTSTIDASGMSVGPASV--ISGL 3401  
 QY 2006 AAGEV--SAISTDAVNGS-----QLYKATQGIANTNELDRIHONENKANAGISSAM 2056  
 Db 3402 ADGTVTISATITTDGAGNSQTHNVQVNTAAVSLSVTISGDNL-----NAAEAG--SAL 3455  
 QY 2057 AMASMPQAYIPGRSMVTGAT--HNGOGAVAVGLSKLSNCGM 2098  
 Db 3456 TLSC-----TGTFATGVYVTLNKGYSAT---IQSGSN 3489

RESULT 10  
 A41477  
 190K surface antigen precursor - Rickettsia rickettsii  
 C:Species: Rickettsia rickettsii  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Oct-1999  
 C:Accession: A41477  
 R:Anderson, B. E.; McDonald, G. A.; Jones, D. C.; Regnery, R. L.  
 Infect. Immun. 58, 2760-2769, 1990  
 A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated,  
 A:Reference number: A41477; PMID:90354033  
 A:Accession: A41477  
 A:Molecule type: DNA  
 A:Residues: 1-2249 <AND>  
 A:Cross-references: GB:U31227; NID:g152465; PIDN:AAA26380.1; PID:g152466  
 A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi  
 C:Keywords: surface; antigen; tandem repeat  
 F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 5.9%; Score 630.5; DB 2; Length 2249;  
 Best Local Similarity 23.2%; Pred. No. 1,2e-16;  
 Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

OY	180	LADINLOKRYRTAOG-----HASTAV-----GAMSYAG-----HFSNAFTR- 218
Db	1	MANISPRLEFKAIQOGGLKALFTTSTAIMLSSSALGVATGVATINNAAFSNVGN 60
OY	219	-----ATAEGT-----YSLAVGLTPTAKAASSIIVGNSQAQIGPAAAVGSGT 261
Db	61	WNETIAGVANGRPAGGPONNAFTYGGDYVTYTAADABRIKAIN-----VAGTT 110
OY	262	QVIN-----RG-----IALFGSOVLKONDVNAANRAVAPDDNOPIDNRK 305
Db	111	PVGINTIQTNTVGSIIITKGLPVTLLMGKSLTLGNNAVNAHGFADPNY----- 163
OY	306	ATFNKATDVFSIGNSNGDSIRKIIYNGASADTDVANAQLEAVRLANROITEKD 365
Db	164	-----TGIGNIALGANAAL-----TIGSAAPSKITLIGNID-----GGIITVKT 205
OY	366	DSNNRVEKGLKLTITTGAGOTSALTDHNIYGVOGDDIKYOLAE--TLTSLMVT-- 420
Db	206	AAINGTIGNTMALATYVAVGATITLG-----GAVIKATTTKLTMAASVLTITNANA 261
OY	421	-ENLTANEKVY---GKRLTPTDKIGFTDMNGIDESKPYLDKDTGIHAGQOKITKLA 475
Db	262	IDNTTGGDNGVLTNMGALSQVTDGIGNTSLATISVG-----AGPATLGAVIKAT 314
OY	476	GVVDDDAATYGOGLKRYNQ-----ASALQTFYVKVKDKNGDADSKIIYV----G 523
Db	315	TKLTDAASAVKFTNPVVYTGALIDNTGNANNGIYTFEGNST--VTGVNGTINLATYV 373
OY	524	KNNKPDG--TOVNTFKLKGENGVDVETNGVTFGL-----NONGLT--VGNSTLNDG 575
Db	374	LLOYGGVYKATINIL--IDNASAVTFTNPVVYTGALIDNTGNANNGIYTFEGNSTVIGD- 430
OY	576	LSVKNTSNKQIOVG-----ADGTFPTDISNKPAGIENRTI 614
Db	431	-IGNTMALATVNVGAGATATLGAVIKATTTKLTMAASVLTITNANA 485
OY	615	TRODGIERNNTGSIIDANKPRLPTPTGINAGKELTNVQASINPATNGOLDPMRNLSTANT 674
Db	486	GGDVGAVGLTNLNGALS-----QVTGIGNTSLATISVGAGPATLGAV-----IKATTT 534
OY	675	EKSASATIKOLYMLSOVPLTFAGDPGPVNTKLGELIKVKGKGTJADTLKRN----IG 730
Db	535	KLTDAASVK---FTNPVVYTGALIDNTGNAN--NGIYTFEGNSTVYTDGIGNTSLATIS 588
OY	731	VVADSTNLSVLKAKTLLSDLAIVNTKTLTASDKV--TVDS--GNNTAKLO--NGDITF 783
Db	589	VGAGPATLGAVIKATTTKLTMAASVLTITNANA 647
OY	784	SKONTGATPATNS-KITGVGVLKFTDMNGIALDGTYYITKDKVYGRAKODGSLDSKPYLD 842
Db	648	--QVTGIDGIGNTSLATISV--GAGPATLGAVIKATTTKLTMAASVK-----FT 693
OY	843	KDKLKGVEVEITTINGIN-----AGKALITG-LSNTITLDTAT--NATTGHVYOLGIV--DST 892
Db	694	NPVYVTAIDSTGANNGIYTFEGNSTVYTDGIGNTALATYVAVGATITLGAVIKATTT 753
OY	893	DKTRAASIGDVLNAGFNLKNNGDAKDFVSTYTDVEINGNATTAARVYD--GKASKVAYDV 951
Db	754	KLTMAASVLTITNANA 809
OY	952	NVDGTTIHLGADGNKNOIGVKTITLTLEKIDAKDKAINESVNSGDDKALINAKDIADN 1011
Db	810	SVGAGPATLGAVIKATTTKLT-----MAASVLTITNANA 845
OY	1012	TLAGEIRTKCTADTALOTFOYKVKVENGD--DDNDADITYGKDAKT-----NOVNTLK 1064
Db	846	-LTGVAVNTTGGDNGVGLTNMGALSQVTDGIGNTSLATISVGAGPATITLGAVIKATTTK 904
OY	1065	LKKGNGIDIONTKDGTVFGINTOSGLKAGNNT--TLNNGL-----SIKNTAGNEQIOY 1117
Db	905	LTNASVLTITNANA 961
OY	1118	GA-----DYKFAKVNNGVYVAG--IDGTRITRIDEIGFAGTN--GS 1155

Db	962	GAGATLGCAGVIAATCTTKLLDASAVKFTNPVYVGAIDNTGNANNGIYTFNGNSTYTCGN	1021
Qy	1156	LDKSKPHLSKDGINAGGKTTTIOSELIQNSNDVATGKIYDLTELENKISSIAKTAQ	1215
Db	1022	VGNITNA-LAAVNVAG-LLIQGVGVVAMNINLTD-----NASAVFT	1062
Qy	1216	NSLHEFVADQGNNTVSNPYSYDTSKTSDIYFAGENGIT-----TYKNKCV	1266
Db	1063	NPVYVGAIDNTGN-----ANNGIYTFNGNSTYTCGNVNTNALATYVNG-----	1106
Qy	1267	RVGIDQTKG-----LTPPKLVG-----NNNGIYDISQNGQNTIT	1303
Db	1107	#AGLLQYQGVGVAMNINLTDNMSAVTFNPVYVGAIDNTGNANNGIY--TFNGNSTYT	1163
Qy	1304	G-LSNTLANTYNDKGSYRTTEGKIIRKDEKTRAASIVD-----VLSAGFNIOGGEAVD	1357
Db	1164	GDIGNTALATVNVAGITLQAGSGL-----AANNIDFGARSTLEFNGPLDGGGKAIP	1216
Qy	1358	F-----VSTYDTVFAQGNMTAKVYDDIITSKTSKVY	1399
Db	1217	YEFKGAIANGNMILANYTKLLASHLLTGTVAEINIGAGNLETTIDASGD-----VT	1265
Qy	1391	DVANVDPTIEVKQKICGVKTTTTLSTGTGANKFPALSNQATGDALVKASDIVA-----	1444
Db	1270	ILNAQNIINFRARDSVLYSLM--IGVGNLIL-----ADLVAPGADEGV	1314
Qy	1443	---HLNTLSGDIOTAKGASQANSSAGYVDADGNK---VIYDS--IDNKYQAQNDGT	1491
Db	1315	VFNGGVVGLN-----VGSNVAGTARNIGDGGNKENTLIIYNAVITIDVNLIEGIQVWL	1366
Qy	1492	VDKTEKVAK-----DKVAQAQTPDGL--AQMNKVSYINKQVYDA	1533
Db	1369	INKNADPTSSIAFNAGAIQIDNATYITIDANGMNLNIPANIOFHAHADQVLQNSSGNDR	1422
Qy	1532	NKKQGINEDNAFVYGLEKASDARKTKMAATVGDLMANV-AQPLFPAGDGTGT---AKKL	1587
Db	1429	TIITGANID-----PDNDDEGIY--LMSVTRAGKLTITAG--GTFPGGAHL	1477
Qy	1588	GETTLIKG-GQDTNKLITDNNIGVAGTDFYVLAKLITNLMSVYVAGGTRIDDKGSFV	1646
Db	1472	-QTLIFAGAGDCST-----AGTFENTNIVLIT-----	1495
Qy	1647	DSSQAKANPVLANSOLDLGKVIYSVNGCTDPTDAANYQOLNEVNRNLGLGNNG---N	1700
Db	1500	--GQ-----LELG-----ATTANVLENDVAVOLTQGTGNIGFLD	1533
Qy	1704	DNAD-----GNOVNADIKDKPSSGSSNRVIKAGVYLGKGNNDTEKLA---TGQVQ	1754
Db	1532	FNARKGAVTLLNNNVAVAGAVO--NTGTTNGTLL--VLGASNLNRVGMIMAKYAGCN	1588
Qy	1755	VGVDRKDGNAV-GDLSNVVKTQKDGSKKALLATYVNAAGOTNYLTNNPABAIDRINEQIR	1811
Db	1586	VTIAKKGKVKIGELQGVGTM-----LTLPAHFNLTG-----SINKTGQALK	1628
Qy	1814	FFHYVNDQNEPVOVQGRNGIDSSASGKHSVAIGFOAKADEAANAIGROFOAGNOSTAID	1877
Db	1629	LNFNMGVGSVCV-----GTAAHSVDDITTAGTSPASSV	166
Qy	1874	NAOAT-----GPOSTA---IGGNVAKKHSAGALGDPSTYKADNSYSVGNNOPTDATORD	1924
Db	1664	NAKGATATLGTTSPANTFTNTGAVTLAK-----GSTSPAKRVAT	1700
Qy	1927	VFGVGNNTIYESVVALGSNSAISACTHAGTOAKKSDGTAGTTTTAGA---TGVYK	198
Db	1705	SP-VANSNTIIFNSLAFNSR--ITG-----GTLILIGANOVYTYGT--GS	1744
Qy	1982	FAGOTAVGASVGA--SGAERRIQNAAGEVSATSDVAVNGSOLYKATOGIAMATV	2035
Db	1747	FTDILLTMTTFDGAAGSGGNILIKSGSTLIDLSGVSTLAL-----VYATN	1791





```

Db 1297 L-----LTHTVAG-----TSNVATIDTISANIDTA 1322
QY 743 KLAKTLDADAVTKTLTASDKVTYDSSGNTAKLO-----NGD-----LFEFQNTGA 790
Db 1323 FVA-----GAAATLTLPVNGAVADGTNOVALVEDANGNPITGAAYVFSSAN-CA 1375
QY 791 TPATNSKTTIGVDLKEFT-----DNNGIALDGTYYITKDKVGFAPKODGSLDKSPYLD 842
Db 1376 TILSTMTNGVGVASTFLTHVYAGTSNVVATIGSTEHDIAFAAGAVATITLAPV-- 1433
QY 843 KOKLAKGEV--ITTINGIA-----GKAIITGLSNTLDATNATGHTVQLGIVDSTDKT 895
Db 1434 -----NGAADVGNYSVQAVVSDSDGNNAVGTATVFSSA-NATAQITTVIG-ITGAGGI 1486
QY 896 RAASIGDVINAGFNKLNNDGADKDEFTVDYFINGNATATKATYDGGKASKAYVNDVG 955
Db 1487 ARTLTINIVAGTSNV-----VATIDTV--NANIDTVA--GELENTV--VSTIN 1530
QY 956 TTHLHGADGNKNOIGKTTTLTKDADKKAIFSVNSGDDKA-LINAKDIADNLTLA 1014
Db 1531 NNALAGADTNIVEAFV-----TDRENGVANOGLIGTGASIVSGSTVTTNLD-- 1580
QY 1015 GEIR-----NTGRTADL-----OTFOYKVKV-----ENGDDND 1045
Db 1581 GRVRSATHTVAGSSNTVIAISGAHOGYARVTVADVSTAQLKLTSLDNOJLANKAGNI 1640
QY 1046 ADITVYKAKTNOV--NTLKLGKNGKIDIOINKDGTVEFT-----NTOSGL----- 1091
Db 1641 AQAALV--DAHDLNLANOSVFRALDNGAVIESOGDSASAGVILKRFNNLTLAGMTVAT 1698
QY 1092 --KAGNNTLNNGL-----SIKNTAGNEQIOVGADGVYFAKANNVVG- 1133
Db 1699 LDTGQTELETHFPAKASTEMTWMKMAVANNIDTNEVOVLVTDVGNAINCAVNL 1758
QY 1134 -----ACIDG--TTRITRDEITGAGTNGSLDKSPHLKDGINA---GCKK 1174
Db 1759 TSNSGNITPNSVTGSDGTATATLHLAGSLPINARDQ---VSKT-INATFIADAS 1813
QY 1175 ITNIOGSEIAONSNDPAVTGKIYD-----LKTLEKNISSTAKAQSLEHF 1221
Db 1814 TQOILAGDFIYNDVOVANGAVNAVQARVDSYGNPIKDOVEVEVLSNNGITQELDVT 1873
QY 1222 SVADEGNNFTVSNPXS-----SYDTSKTSDVIT-----FAGENG 1256
Db 1874 SV--EGVAVTEFTNLGITTNTATVSSGSSRNIDTFTFADVTANHLAASDLMTVDA 1931
QY 1257 ITTKVAKGVYRVIDOTKLTLPKLTGVGNNGKGIYIDSQNGQNTITGLSNTLANVTDK 1316
Db 1932 VADNLDKNEVHARVTDKAGNVLSGQTVFTSGNGAATTVNGISDGDITKATLHTLAG 1991
QY 1317 GSVRTTEGKIITKDEDEKTRAA--SIYDVLASGFNLQNGEAVDFSTYDT--VNFADGN 1371
Db 1992 TSVYVARVGNRQSKDTFIADRTATIRASDLITRRNALADGAVTAANARYIVTDAGN 2051
QY 1372 AFTAAVYDDTSTKTSKVYVDVNDVDTTEVKDKLGVKTTTLTSTGTANK---FALS 1426
Db 2052 PVPMSFVGYTSDNGALFPTSGMTDS-----GTFSTFTHTTAGISCTAIVTMG 2103
QY 1427 NGATDADLVKASDIYAHNLTLSGDIOFAGKASQANS-----AGYVADGN---KYI 1475
Db 2104 ISQTDAAFIARSTAHVS---ELIVVKNSTLANSRNIYQAHIKAHGHWVTGMVNV 2159
QY 1476 YDSTNNKYVQAKNDGTVDKTEKAVADKIVAQOTPDGTLAQMNVKSVINKEDVNDANKQ 1535
Db 2160 FSATENVTILTANTVT--NSQCYAENTLRHNAPTVSAITAIV-----ATLIV 2204
QY 1536 GINEDNAEYKGL-----EKAASD---NKTNAAVTVGDLNAAVQPLTFAGDTG 1581
Db 2205 GLEDEVREYAGAGRIELEFRINDGAVADGICITNREARVYDSD-NLVPSNVVFSADNG 2263
QY 1582 -----TTAKLGET-LTIKGGQDITKLDNNTGVAAGT-----DGFYVYKLAKD--- 1624

```

```

Db 2264 GOLVONDVOTDALGSAYTVSNINTGVTKVTVADGVASATTTFTIADRDATLVTRFL 2323
QY 1625 LTN-----LNSVNAAGTKIDDKGVSEFVSSGAKNNTVYLANGLDGKV 1670
Db 2324 LTHDNAANGVENRVLHLHLDADNDSVSGVEVFSAITNGASINASAITDINGPAIG--V 2381
QY 1671 ISNVKGTGKTDDANVOOLNEVRNLGLGNAGNPN-----ADGQOVNIADIKKDPNS 1722
Db 2382 LNTLSGSDVTVLVLP-----CGTESLVTPOFIADINTANIA----- 2421
QY 1723 GSSNRYVYKAGTVLGKGNNDTEKLAGVGVGVNDKXGANNGLSNVWK----- 1773
Db 2422 --TGDFTVIDGAVANSYDANEVBARVY-----DMQGNIAIGSYVFESSONGATITTS 2472
QY 1774 --TQKDSKRALAPYNAAGOTNLN-----NPAE 1802
Db 2473 GITVDGKASAKL--THIKAGESILARLSRPMATVHTLMPFIADSTATLOLFENP-- 2529
QY 1803 AIDRINEGIRFE--HYNDGNQEPVYVQGRNGIDSSASGRHSAIGFOAKADEBAVAIG 1859
Db 2530 -IPIIADGVMOFFLIGRFEDANQNPV--GGQVAFSATNEVTLIESNGSISTEGSVLLS 2586
QY 1860 -RQTOAGNOSTAIDNQAQIGDOSIAIGTGNVYAKHSGAIG---DPSTVKAADNSYSV 1913
Db 2587 VTSTQAGVHPI-----TGLVSNNTYDTPFGAFTANKNTAOLSTLMVY 2629
QY 1914 GNNQOFDQTQDVEYGNNTITVESNVALGSASISAGTHGTOKK-----SD----- 1964
Db 2630 DNMLADGVTINOVRA---HYVDSTGNSVA---DMAVTFANRAQLSKATVYLLDNNGA 2683
QY 1965 -----GTAGTTTGTAGATGVYKGFAGOTAVAGV--SVGASGERRIQ- 2003
Db 2684 VNTLTNSLVGTVYATLGTATGTFPLVDYFTAGPLATLTLVTVYNNAFADNSATNTVQA 2743
QY 2004 -----NVAAGEVSAISTDAVNSQLYKATQGIANATNELDRIHONENKANAGISSAM 2056
Db 2744 TLKDVSGNPVIGEVAFRA--ASNCAITATDGVGSNAG---IVLATLTNGTAGSTVY 2797
QY 2057 A-----MASMPQAYIIPGRSM-VTGGIATFHNGQGAVA---VGLSKLSDNGQWFKINGS 2106
Db 2798 ATTELTETTTDTTIANKNLDVYVNGTTENGDNAGFPPTGFGVAT-----FKVNSGG 2848
QY 2107 D 2107
Db 2849 D 2849

```

## RESULT 13

F90073  
 hypochelical protein SA2447 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: F90073  
 R:Kuroda, M.; Ohka, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
 ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; PMID:21311952; PMID:11418146  
 A:Accession: F90073  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2271 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BAR43752.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA2447

Query Match 5.5%; Score 590.5; DB 2; Length 2271;  
 Best Local Similarity 17.0%; Pred. No. 4.1e-15;  
 Matches 385; Conservative 435; Mismatches 973; Indels 469; Gaps 78;

24 KSHSTGGSCATGVGVSVRLSPARIALAVLIGATLNN-----GSAYAGIG 70  
267 KOYFTTSGNATYDOSTGIVTLTFODAYSOKAATITGTRIDSKSHFEGSKVALNGKYYGHC 326  
71 ISEADGKG-----GANABGKSIAGIDIAOAGSOSIAGDNKIVHNSNNNN 120  
327 -----NGGDIGFAFSPGLGETGLNGAVALGIGLSNMFGEFL-----DHYHNT-SKPN 375  
121 GAKASGNESTIAGDVLASGHASIAISDLYLKEFYVOQISELLPIITROKALINDIYOL 180  
376 AAKANAPDSNVAGGAFYTTDYSYGAATYSSSTADNAAKL-----WYOPNNFOD 430  
181 ADTNOYKRYTHAGHASTAVGAMSYAKGHPSNAFGTRATAGEYSLAVGLTATKAASS 240  
431 FDIY-----NGDKYMYKYGAGQWTRNISPMIAKSGITNLSLSTASTGGA-- 478  
241 IAVGSMOAIQFAA-----TAVGSGTOVNLNGIAL-----GFGSOVLKONDVANAAN 288  
479 -----TNLOQVOFGTFEYESAVTQVRVYDVTGKOLIPKTYSGNVQVYVITIDNOQSALT 534  
289 VRATAPDNDPIDNRYKATPEKNGATDVFSIGNSNGNDSTIR-KIINVGASADTDAVNA 347  
535 AKG-----NYTSVDSASYST-----NDNTKIVKKTNAG----- 564  
348 QLRBAVRLANROIFPKGDDSNRVEKGLKTLITGGAQOSAL-----TDHNIG 396  
565 -----QSVTYFTDV-----KAPITVGNQITLEVKTNNPVYLTITDNGTG 605  
397 VWONG-DGLKVOALETLSLMTTENILANKEVYVGTBLTDTIGFTDMNGIDSKP 455  
606 TVTMTVGLPGLSYDSATNSIIGTPKIGOSTVTVSTQDANKSTTTFIINVVDITAP 665  
456 YL-----BDOTGHAGOKITKILTAGVDDDAITYGOLKKVQTA 495  
666 TVTPIGDSSEVYSPISPIKATODNSGNNAVNTVGLPSGLTDP-----STNNTI 716  
496 ES-----ALQFTFKYKDKNGNDANDSKITVGNKNNKPDG-----QVNTLKG 540  
717 SGTPTNIGTSTIISTVSDASGNKTTTTPKYEVYTRNSMSDSVTSIGTSQOSVSTSKADS 776  
541 ENGVDVTEITNGV-----TEGLNONGILVGNSTLNDGLSVKNTNSNKOIOVAGADITF 596  
777 QS-----ASTSGSIYSTASTSKTSVSLSDVSASAKSLSTSSNS-----VSSSTS 826  
597 TDISNKR-----GAGIENTRITRDGIGFANNTGSLDANKPRLTPIGAINAGKELTANQ 651  
827 TSLVNSQSVSSSMGVSYSKSTSLSDISNSNST-----EKSESLSTSTSLSTSTSL 880  
652 SAINPATNGGOLDPMNRSTANTKESGAATIKDLYN-----LSQVPLPAGDTGPNVTKL 708  
881 DLSKMSIS-GSLSKSOSLSTISIGSSSTSLASLSTSTSLSTSLSTSLSTSLSTSL 938  
709 GEILKVGKGTADDLTKNNIGVAVDSTNSLTVKLAKTSLDLDVAN-----TKTLTASD 763  
939 NSIANSQASSTKSDOSTISL-----STPSDKMSSTSESLSDSTSGVSGSLSTASQ 995  
764 KYTVDSGN-NTAKLQNDLTFKQNTGATPATNSKITGVDGLKFTDNNGIA-----IDGT 817  
996 SVSTSTSDMSSTSEISVDSIST-----GSLASDSKMSVSSSMSTSGSSTSESLSDQ 1051  
818 TYTRKDKVGA-----KODSGLDKSPYLDKOKLKYGEVITNGINAGKAITGSLNLTLD 874  
1052 STSDSDKSLSTSLSTSGSSTSTSTSTSTASVRISESOSTSGSASOSDSMSISTSTSPD 1109  
875 ATNATTHGVTLGIVDSTDKTRAASIGDVLNAGFLKNNNGKAMPFVSTYDVTVDINGAT 934  
1110 ST-----SDSKASSTASSESTISQASSTSTSGS-----VSTSTSLSTNSERT 1151  
935 TAKTYDCKASKAVADVNDGTTIHLGADGNKNOIGYKTTTLTKTDAKGKAINESV-- 992  
1152 STVS-----DSTSLSTSESDISSESTSTSTSTSTSEISAIASESTISLSESN 1197  
993 NSGDOKALINAKDIADNLNTLAGELIRNTKGTADALQTFQYKAYKENGDDNDNDADITTVG 1052

1198 STSDSESGASAPLSSESLSESTSESTSESVSSSTSTSTSLSDSTSPSSGSTSLSNSTGC 1257  
1053 KOAKTNOVNTLKLKGNKNGIDIQTKDGYTFEGINTQSGLKAGNNTTLNNGLSIKNTAGN 1112  
1258 SASISTSTSTSE-----STPFKESEVSTSLSMSTSTSLSNSTSLSTSLSDSTSDSKS 1310  
1113 EOIOVG-----ADGVKPAKVNNGVAGCIDGTTRTRREIOFACITNGSLDKSKHLKXDGIN 1169  
1311 DSIJSTMSSTSDISTSKSDSISTSTSLSGSTSESD-----STSSSEKSD----- 1357  
1170 AGGKLTITNIOGSEIAONSNDVAGKIDYDKTELEKISTAKTADONSHEPSVADEQCN 1229  
1358 -----STMSISMOSSTSGSTSTSTSTSLSDSTSLSLSLASMSMOSGVSNSASQASAN 1411  
1230 NFTVS-----NPYSSTYKTSKSDVITTFAGENGITTKVKNGVYVAGIDQTKGLTTPK 1280  
1412 STSTSTSESDOSTSTYTSOSTSOSTSTSTSLSD-STSIKSTSGSSTSTSLSGS 1470  
1261 LTVGNNGKGIIVDSQONQITITGLSNTLANVYNDKGSVRTTEQKIIDDEKTRASIY 1340  
1471 ESESDSOSISTSTSESTSEASTSLSDSTSTNSGASSTSTSLNSASASESDSSSTSL 1530  
1341 DVLAFGNLQNGEAVDFSTYVNFADGNATYAKVYDDTISKSKVYDVAVDVTIE 1400  
1531 DSTSASMOSSESOSSTASLSDSLSTSTSNKSTIASLSTVSTSE-----SGSTSESTSE 1587  
1401 VKDKLGKVTTLTSTGTGANKFALSNOATGALVAKASDIAVHLNTLSGDIOTAKGASQA 1460  
1588 SDSTSTSLSDOSTSTSTSTASGASSTSTSDSRSTSTSTSMRTSTSDOSMSISTST 1647  
1461 NSSAGYVDADGNKVIYDSTDNKRYQAKNDGYDKTEVAKKDLVQAQPTDGLQMANVK 1520  
1648 STS-----MDST--SLSDSVSDSTSDSTSTSTSTSTSTSTSTSTSTSTSTSTST 1694  
1521 SVIN-----KEOVNANKKOGIENDENAFYKLEKASDNKTKNAVATGDLNAVAQ 1571  
1695 EVMSASISDSOSKESVNDV-----ESVSESN-----SESDSKMSGSTSVSDSGLS 1742  
1572 TPLTFAGDTGTAKK--LGETLTIKGGOTDTNKLTDNNIGVAGTGTFTVKLANDLNL 1628  
1743 -----VSTSLRKSSESVSESSLSGSGS-----VSTSDSSSLSTVSTSLRS 1786  
1629 NSYNAGGTIKIDKGVAFVDSGAKANTPYLSANGDIDGKYVSNVGRKTOXTDANVQ 1688  
1787 ESVSESDSLSDSK-----STSGSTSTSTSGSLSTSTSLSGS--ESVSESTSLSDSTSMED 1839  
1689 LNEVRNLLGNGNNDNADGN-OVNADIKKDPNSGSSNRVYKAGVYLGKGNNDTEK 1747  
1840 STSTSDSL--SGSISLSSTSTSLSDSLSDSKSLSSOS--MGSGSTSTSVSDSGS 1894  
1748 LATGVQV-----GVYDKGNANGDLSNVVWTKQDKSKKALLATYNAAGQNTYLTNDA 1801  
1895 SSTRNSQFDSMSIASSESDMSSTSDSSNI-----SGSNSTSTSLSTPSMGASVSTST-- 1949  
1802 EADIRINEGIRFHHNDQEPVVOGRNCIDSSAGKHSVALGFPKAKDGEAVALIGQ 1861  
1950 -----SLSDS-----ISGSTSVSDSSSTSTSTST----- 1973  
1862 TOAGNOSIAGDNAQTPDOSIAGTGNVYAGKHSALIDPSTVKADNGSVGNNNQFDD 1921  
1974 --SDMSQSGSTSTASGSLSTSTSTSMASASTS-----SQGSTSVSTSLSTSD 2025  
1922 ATQTDVFGVGNNT-TVYESNSVALGNSAISAGTHAGTQAKKSDGATTTTATGATGVK 1980  
2026 STSISISQSTVESESTSTSTSDSESLSTSDSTSTSTSTSGSTSTSTSESLST 2085  
1981 GFAGQTAGV-----ASVAGSAGERRIQVAAQGVASNT-----DAVNSQ-- 2022  
2086 SGGSGSTSVSDSTSMSESDSTVSMQDKSDSTSTSDSESVSTSTSTSTSTSTSTSESL 2145  
2023 --LYKATOGIANT-----NEL-----DRIHONENKANAGIS--AMAMASMP 2062

Db 2146 STSMGSGSISDSTSTSMGSGSTSTSESNMHPDSMSMHHTSTSTSLSEATTSTSES 2205  
 QY 2063 QAYIPGRSMVTGATIHNGGAVAGLSKSLNGQWVKING 2104  
 Db 2206 QSTLSATSEVT---KHNGTPOASE--KRLPDTGST--KONG 2240

## RESULT 14

AB3528

extracellular serine proteinase (EC 3.4.21.-) [Imported] - Brucella melitensis (strain 1  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AB3528  
 R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: AB3528  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-2554 <KUR>  
 A:Cross-references: GB:AE008918; PIDN:AAU53389.1; PID:917984282; GSFDB:GN00191  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BMEI10148  
 A:Map position: II

C:Keywords: hydrolase; serine proteinase  
 Query Match 5.3%; Score 568.5; DB 2; Length 2554;  
 Best Local Similarity 21.4%; Pred. No. 3.4e-14;  
 Matches 492; Conservative 271; Mismatches 925; Indels 607; Gaps 104;

QY 202 GANSTAKGFSNAFGRATAEGLTSLAVGLTATAKAASSIANGNAQALGFAPAT----- 256  
 Db 20 GAIQFTR---AGAGSVTLTGQNTYSGVTLTSGILTVTOMANAGASGIG--OSTADPANTL 75  
 QY 257 ---VGSSTOYNLRGIALGFGSOVLQKNDVNANRARAAPPDNDPDRK-- 305  
 Db 76 MLESGFRTTGS--VTTBRGFLVNGPARIETVGTSSNL--AFSGLVTSDDAGFEK 132  
 QY 306 ---ATFKGATDVFSIGNSNGDSIRKRIIVG---AGSADTAVVNAQLKEAVRL-- 355  
 Db 133 GAGTLFLFNGSDHIGATVSGTLAVSTLADGQVSSLGKSGSDATNLLAGALNTYG 192  
 QY 356 ---ARQITFKGDSNNRVEKGLKTLITGAGQTSALTDHINICVYQGDGLXQVLAET 411  
 Db 193 STTSSDRSFTLGAAGSICVANA--GTTLSMSG--TAVGTD--GLTKLGDG----- 237  
 QY 412 LSLKAVTENTLANEKVTVGKTRLTPTDKI---GFTNDMNGIDSKPYLDDGTGHIAGG 467  
 Db 238 --TLISGTNTYTGNTAVNAGVLRAGSAQAFPSGLMTYVNG-----ASLELGG 284  
 QY 468 QKIT---KLTAQVDDDAATYQGLKKNVQTAESALQTFVTKVVDKNGNDANDSKITTYGK 524  
 Db 285 YDITVSGLLGAGTVLDGNTL-----TSSGSAANSFTKILTGIGFRTGSGTQTLISG 337  
 QY 525 NKKPDTQVNTLKLKENGENDVDTTETNGVTFGLNQN--NGLTVGSLTANGLSVKKN 582  
 Db 338 CNS-DYTGKTTTASNGTSLVDCLKNGQASSIGASSNADPNLVNNGTSLYGTNTV--TTD 395  
 QY 583 SNKOIOVAGDGIPTDINSKPGAGIENTTRITRDGIFPANNVTSV---DANKPRLPTPG 639  
 Db 396 RGTIDGCGALSVTAATLTFSSG---QYVGTALQKRTGTGLVLMNSYSR--GGTS 449  
 QY 640 INAGKELTNVOSALNPATNGOLDPMNRLSTANTEKSSAATIKDL--YNLSQVPLPFA 697  
 Db 450 VDAQ-----TLRAGSSGARGGSGMSLSN-----AAGAILDIDGFTSVTSLSGG 493  
 QY 698 GDTGPVNTKKGILKVGKGTADLT-----KNNIGV-----VAOSTNSLTV 742  
 Db 494 GALGQVVALG--GATLTITSSGNSNGTSYTGAITGIGNFVANGTORLTCGASSYSQSTTI 552

QY 743 K---LAKTSLDDHVAVTKTLTASDKVT--VDSGNNTAKLQNGDLTFESKONTGATPATNSK 797  
 Db 553 NGVLEWESCLADGGSVSSISMSADADNLVINGVLTAKYTGSGDSTDRQFLAGSGNSIE 612  
 QY 798 TIVGDLKFTDNNGI---ALDGTYYITKDKVGRAKQDGSIDSKSPYLDKDLKVEYEIT 854  
 Db 613 SEGTGAILFTSMAAVTFPAANTAGTLTLTLAGTNTDNERG-----AQLT 657  
 QY 855 TNGINAGKMTIGLSNT-----LTDAFNATGTVHVL--GIYDSTDK---TRAASIGDV 903  
 Db 658 NN-----GSGTISLTKIDTGTWFLTNSDSTYTG--VTKINGVLSVDKLAGVLAASSIGAS 711  
 QY 904 LNAENIKKNGDADPEFSTYDVDFINGNATKATKATVYDGRASVAVDNDGTTIHLTGA 963  
 Db 712 SSASNL-----IIGNDSTLRKLTGDTTDLRLFLASGLTYIESGGS 753  
 QY 964 DG---NKNQIGVKTTLTLTKDAGDKAINEFVNSGDKALINAKDIDLNLAGEIRN 1019  
 Db 754 GAIVFTDYGVALADNNQARTIALGK-----NTGD-----NRLASIGD 793  
 QY 1020 EKGADTALQTFQYKVKYKENGDDNDADTIVGKDAKTQNVNLT--LKRGNGLDIQTNK 1077  
 Db 794 A-GTCKITLA-----KDDGTGYLTGNNITGPTNINKLKLKIGNG-----GT 835  
 QY 1078 DGTVTFGI--NTQSLKAGNNTLNNGLSIKNTAGNEQIOGVADGVKFAKVNNGVYAG-- 1135  
 Db 836 TGSLSLDIVYDGLIFNRSDTLNYYGL---ISGAGFVYQSGSGTILGANSYTGATS 891  
 QY 1136 -IDGTRITRDEIFAG---TNGSLDKSKPHSKDSKINAGGKKTINIGSGEIAONSND 1190  
 Db 892 VSAGTLLVNGDOSANTQTSVANSI---LEGSSIIIGGNVYVD--GALAPSGNGA 942  
 QY 1191 ---VTGKIKYDKTELEKNIS---STAKTAONSL-- 1218  
 Db 943 GTLLTNGSLAISASIIISMOLGQAGVAGALNDL--IEVKNLITDGLVDVAEFAAGSYCP 1001  
 QY 1219 ---HEFSVADEQGNFTVSNPSSYDTSKTSVDVITPAG-----ENGITTKVNGK 1264  
 Db 1002 GIYRLINTGSLTDNGLDIGMLPAGAGAIQTAVAGQVNLLAGTNEFMFGDVGDPFNSA 1061  
 QY 1265 VVRGIDQTKLTPKLTGVNN-----GKIVIDSONQONT 1301  
 Db 1062 V-----DGGNGTWQSSGNNTNMTDATGMINASYSDGAPALFTGATVYDINSIGQVK 1114  
 QY 1302 ITGLSNTLAN--VTNDKGSVTFEQGKIIDEKTRA-----ASIVYLSAGFNLOGNGE 1354  
 Db 1115 AEGMQFAIDSAVYGDK--LELTGPQSTIRVGDETGAAYIAVNTSVLGNQGLE--KTD 1171  
 QY 1355 AVDEVSTYDVNPDG---NATTAKVITYDT--SKTSKVYDVAVVDTTIEVKDKKLGK 1409  
 Db 1172 AGTIVLT--GANSYTGTAINGTIRISSDNLGYASSDISPFGGALMTNTANITADRAIIL 1230  
 QY 1410 T-----TTLTSTGTGANKFALSNOATGDLVYASDIVAHLNLTASGIDIOAKGASQ 1459  
 Db 1231 TGAGTLTLTASTTSLSGISGTLKSGTGTL--SGTAH---TGGTITTTACTLO 1284  
 QY 1460 ANSSAGYVADAGNKV---IYDSTDKKYYQAKNDGTVDYTKE-----VAKDKLVAQA 1507  
 Db 1285 IGNKGTDSIDGNIVNNCALVFDRAGLATYAGSISGIGTLKKNSSLTLMWGTSTYGET 1344  
 QY 1508 QTPDGTAAQ--NVKSVIN-----KEQVNDANKKOGINEDNAFVGLKKAASDKTKN 1358  
 Db 1345 TVSAGTTLAQAGQIKGASLTVDGAEVLIDGSGSOFATGAGASVYG-----T 1393  
 QY 1559 AAVYVGDINAVAQPLTFAGGTGTAKKLGTLTIKGGQDPTNKLDDNNIGV--VAGT-- 1614  
 Db 1394 GTVYVRDGTGASPSDLTSTNATGINS---TIVVAG--SGSOMTGTGATGELAGTAT 1445  
 QY 1615 ---DGFTV-----KLAIDLTLNLSVNAAGTK-----IDDKGVS 1644  
 Db 1446 VDILDGTMISSGASVYFGGGLPMDATGQVYISGAGSOWTIANLVARSGSITYDDGCV 1505



Db	144	LNG-----	NNADANHGEGAPADNYTGLNLI	ALGANAALLI	IQSAAKPIITLAGNIN	195
Qy	116	NNANIGAKKASGNEISALIGC-	DVLASGHASIALGSDDL	YLAKTVOQISELLP	IIRGOKAL	174
Db	196	GGGIIYTKDMDAINGITGNIN	LATVNAVAGATLEGAIK	TKTTTKLINAASVL-----		249
Qy	175	NDIYQIADNTMLQKRYRTHAOGH	SHASTVGVAMSAKGFHSNAFSTR	---ATAGEYSLAVG		230
Db	250	-----	TLINNAVILTGADIDNTGTV	DVNAVGLINLNGALISQYTG		285
Qy	231	LTATAKAASSIAVGSNAQAIG	FAPATVAGSGTOVINR	GIALGFSQVLOK---DNDVAA		287
Db	286	NIGNTNALATISVAGAKATL	IGA--VIAKATTKLIDN	ASAVFTNPVVYTGADIDNTGNA-		342
Qy	288	NVRARVADDQNPIDNRYKAT	FEKNGATDVSIGNSGND	IRKTIINVGASAD-----		340
Db	343	-----	NNGIYFTGDSIVTGNIGITNA	-----LATISVAGAKATL	GGALIK	383
Qy	341	-----	TDAVNAVOLKEAVELA-----	NROITFEKDDSN-----	RYEKG	374
Db	384	ATTTKLIDNMSAVTFNPVVY	TGAIDNTGNANNGIYFTG	DSIVGNIGNTALATISVG		443
Qy	375	LGLTLITTG---AQISAL	TDHNIGVYQNGDLKYVLA	ETLISLKMV-----TENLTAN		426
Db	444	AGRA-TLGGAI	IKATTKLIDNMSAV-----	TFNPVVYTGADIDNTGNAN-N		488
Qy	427	EKVAVGKTRLTTRKIGETN	MNGIDSKPYLDKDTIHAG--	-GOKITTKLAGVDDDA		482
Db	489	GIYFTGDSIVTGNIGNTNA	LATI-----	SVGAKRATLGGAI	IKATTKLIDNA	537
Qy	483	ATYQOLKRVNOT-----	AESALOTFT-VKKVDK	NGDANDASKIITVYKNNKP-DCTQ		532
Db	538	SAVFNPVVYTGADIDNTG	NANNGIYFTGDSIVTGNIG	NTALATISVAGAKATLGA	IAI	597
Qy	533	VNTLKLGENGVDTTETNCTY	FFGL-----NONNGLT--	VENSULINDGSLVYKTNENK		585
Db	598	EKATTKLIDNASAVFTNP	VVYTGADIDNTGNANNGI	YFTFGDSIVTG---NIGNTNLA		654
Qy	586	QIOYGA-----DG-----	ITFIDISNRKPGAGIENTR	ITRBDGIGFAPNN		624
Db	655	TVNAGAGATLEGAVIKAT	TKTKLINAASVLTLINVA	AVILGA-IDNTGV--DNVGVNL		711
Qy	625	TGSLD-----	ANKRLLTPGICANGKELIN	VOSAT--NPATINGOL		663
Db	712	NGALSQYVGNIGNTNAL	ATISVAGAKATLGA	VIKATTKLIDNMSAVFTNP	VVYTGAI	771
Qy	664	DEMNRLSTANERKSGSA	ATIKDLNLSQVLPFAG	OTGNVYKRIKLE-----I		711
Db	772	D-----	NTGNA-NGIA-----	TTGGS--TVTGNIGNTNAL	ATVNAVAGL	809
Qy	712	LKVGKGTADDLTKNNIG	AVADSTNSLTVKLAKTL	LDLDAVNTKTLTASDKVT	VDSDG	770
Db	810	LRYVG-----	GVKXSNTIN-----	LTD--NASAVFTNPVVY	TGAI	843
Qy	771	NNTKALONGDLTESKON--	TGATPATNS-KTIEVD	OLKFTDNNGIALDSTYIT	ITDKYGF	827
Db	844	DNTGNANNGIYFTG	DSIVGNIGNTNALATIS	VAGAKATLGGAI	IKATTKLIDNA---	900
Qy	828	AKOGSDLKSPYLDKRLK	VEVEITFTNGINNGK	ALIGLSWTLIDATNATGH	YTOIG	887
Db	901	-----SAVFTNPV-----	VYTGADIDNTGNANNGI	V-----FTGDSIVTG		936

```

Db 1035 ASAGDVTILNAODI--HFRAL-----DSALVLSNLGTVGN---NILLAADLV 1077
Qy 1050 TVGDAKTNQOVNTLKLKGNGLDIOFNKDGTYFEGINTSGKAGN--NTTLNNNLSLK 1107
Db 1078 APBVDGCTVFD---GGVNGLISNVAGAA-----RNIGDVGKKEFTLLIYNAVTT 1128
Qy 1108 NTAGNEOIOVGADGVFAKYNNGVAGIDGTRIRTRDEIFAGTSGSLDKSKPHLSKDQ 1167
Db 1129 D-----DVNLEGIONVLINN---NADFTSSJAFNAGTIOQINDATYITDANNGNLN--- 1175
Qy 1168 INAGGKKTINISGELIAONS--ND-AVTGKIYDLKTELENNKISSAKIAONSLEHFSVA 1224
Db 1176 IPAGNIKFAHAAQILLOHNSGNDRTITLGANIDPDNDGEGIILNSVTA----- 1225
Qy 1225 DEOGNFTVSNPYSSYDTSKTSVDITFEGENGITTVKNKGVVRVIGDQTKLTPEKTVG 1284
Db 1226 ---GKLLTAGGKTPGAKRLDIIY-FKGE-----GDFGTAGTTF----- 1261
Qy 1285 NNNKGIVIDSONGONTITG---LSNLTANVTNDKGSVRTTEOGKLIKDEDKTRAASIVD 1341
Db 1262 --NTTNIVLD-----ITGOLELGATFANYVLEKDAVQLTQTG----- 1296
Qy 1342 VLSAGFNLOGNEAVDFVSTYDTPVNPADGNATTAKYTYIDTSTKSKVYDVAVDDTTIEV 1401
Db 1297 ---NIGG-----FLDENAKNGVTLLNN-----NNVWAGT----- 1322
Qy 1402 KDKKLGVKTTTLTSTGT---GANKFALSNQATGDALVAKSADIVAHNLNTLSGDIQTAKA 1457
Db 1323 ---VKNTGGINNGTLLVIGASNL---NRVNGIAMLKVG-----AGNVTIAKG- 1363
Qy 1458 SQANSSAGYVDADGNKVIVDSTNKKYQAKNDGTVDKTEVAKDKLVAAQOTPDGTLAQM 1517
Db 1364 --GNVAKIGEIGTGTFTL--FLPAHF--KLGSINKT-----GGQALKL 1401
Qy 1518 NYSVINKQOVANDANKKOCINEDNAFVKLEKASDNKTKNAAYVGDILNAVQTPLEFA 1577
Db 1402 N-----FMNG-----GSVGVVGTAAHSV 1420
Qy 1578 GDTGTT-AKKLGTELTIKG---GQDTNKLFDNNIGVAGTDFTVKAKDLT----- 1626
Db 1421 GDITTAGATSPASSVAKGTATLGTTSPAHFTNTGAVTLAKGSTSPAKKVTATSEVA 1480
Qy 1627 NLNSVAG-----GKIDDKGV-----SFVDS-----SGQAKANTPYL 1659
Db 1481 NSATINFGNSLAFNSNITSGTTLILGANQVYTGTSFTDTLTLNTFPDGAASKGNIL 1540
Qy 1660 SANG--LDIGKAVISNVGCTDPTDAANYQULNEVRNLLGLNAGNDNAGNOVINADIK 1717
Db 1541 IKSSTLIDLSG--VSNLALVVTAT-----NEDMN----- 1567
Qy 1718 KDPNCGSSSNRTVIRKAGTVLGKG-----NND-----TEKLANG 1751
Db 1568 ---NISPDTKYTVISAETAGGLKPTPKENVKTTINNDNRFVFTPDASTITLFAEDIAAG 1624
Qy 1752 GVOGVYDKDGNANGDLSNV-----WVKTKDGSKK-----ALATYNAAGOT 1793
Db 1625 ---VIDEDFAGGGLANIPNANIKKSLIEMEDAPNGSDARQAFNFGIMTPIQADAT 1680
Qy 1794 NYLTNN---PARAIDRINEOGI-----RPFHVNDGNQEPVYQGRNGIDSSAS 1837
Db 1681 THLMQDVAKPSDTIAVNNQVVASINSITTAALNARMKVOAGNKGPVSSGDEDMD--- 1736
Qy 1838 GRHSAVIGFOARADGEAAVAIGRTOAGNOSTAIGDNAOATG-----DOSIAIGTG 1888
Db 1737 -----AKFGAMISPFVGNATQKMCNOSTIS-GYKSDTTGCTIGFPGFVSDDLVLGLA 1785
Qy 1889 NVVAGK---HSGAIGDSTYKAD-----NSYSGVNNQFTDA----- 1922
Db 1786 YTRADTDIKLNKNGKDGKAKVESNITSYGLYSVYENLFVEALIASYDNKIRSKSRVYI 1845
Qy 1923 ---TQTDVFGVGNNTITVMSNSVAL-----GNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLLETVGIQTANGKYYKSSSTGQLMAGYTYMSENINILTPLAGLRYSTIKDKSYKERTG 1905

```

```

Qy 1970 TTTAGATGVKGFAGQFATVAVSVAGSAGERRIQNVAAGEVSATSTDAVNSQLYKATQ- 2028
Db 1906 TY---QNLTVKGNKNTFPDGLGAKVS-----SNINVNETVLT-----PELIYANVDY 1949
Qy 2029 GIANATNELDRIHONENKANAGISSAMAMAMPQA 2064
Db 1950 AFKNKVSALDARLO-----GWTAPLPTNSFKQS 1977

```

Search completed: July 30, 2002, 15:37:07  
Job time: 332 sec

Wed Jul 31 07:54:12 2002

us-09-813-214a-9.rpr



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:59 ; Search time 27.4 Seconds  
(without alignments)  
2998.645 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MHHTYKVFKNATGTFMVA.....NGSADTQGHVGAAGVGRHF 2122

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues 105224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	630.5	5.9	2249	1 OMPA_RICRI	P15921 rickettsia
2	550.5	5.1	2021	1 OMPA_RICCN	O52657 rickettsia
3	544	5.1	2003	1 YDBA_ECOLI	P36666 escherichia
4	494	4.6	1656	1 OMPB_RICJA	O06653 r outer mem
5	493.5	4.6	1953	1 BIGA_SALTY	P25927 salmonella
6	480	4.5	1643	1 OMPB_RICRP	O53020 r outer mem
7	478.5	4.5	3591	1 FHAB_BORPE	P12355 bordetella
8	474.5	4.4	1577	1 HLVA_PROMI	P14666 proteus mir
9	456.5	4.3	1655	1 OMPB_RICCN	O9K63 r outer mem
10	444.5	4.2	1902	1 P3P_IACIC	P15292 lactococcus
11	437.5	4.1	1902	1 P1P_IACIC	P16271 lactococcus
12	437.5	4.1	1902	1 P2P_IACIC	O53047 r outer mem
13	435.5	4.0	1902	1 OMPB_RICRI	O02470 lactobacilli
14	433.5	4.0	1645	1 P2P_IACPA	P69689 r outer mem
15	430	4.0	1608	1 HLYA_SERMA	P53320 serratia ma
16	425	4.0	1569	1 YPUA_ECOLI	P52143 escherichia
17	403.5	3.8	1848	1 CHPA_CLOCL	P38058 clostridium
18	398	3.7	2334	1 WAPA_BACSU	O07833 bacillus su
19	380.5	3.6	1325	1 YDEK_ECOLI	P32051 escherichia
20	366	3.4	1300	1 120K_RICRI	P14914 rickettsia
21	360	3.4	1176	1 SLAP_BACSH	P58537 bacillus sp
22	354	3.3	1829	1 FRPC_NEIMC	P55127 neisseria m
23	353	3.3	1829	1 FRPC_NEIMB	O9J1YV neisseria m
24	349.5	3.3	1723	1 PM20_CHLPN	O92812 chlamydia p
25	342	3.2	1025	1 SLAP_CAUCR	P35828 caulobacter
26	339	3.2	1286	1 AIDA_ECOLI	O03155 escherichia
27	338	3.2	1694	1 IGA0_HAETN	P44969 haemophilus
28	329.5	3.1	1567	1 ICEN_XANCT	P18127 xantomonas
29	325.5	3.0	1702	1 IG42_HAETN	P45386 haemophilus
30	325	3.0	1849	1 IG42_HAETN	P45386 haemophilus
31	321.5	3.0	2492	1 TALC_DICDI	P54633 dictyostell
32	321	3.0	1770	1 PMPC_CHLTR	O84419 chlamydia t
33	310.5	2.9			

34	309.5	2.9	918	1 YMOB_CAEBL	P34487 caenorhabdi
35	301.5	2.8	1672	1 ICEA_PANAM	O9PJY2 chlamydia m
36	300.5	2.8	1322	1 ICEA_PANAM	P20469 pantoea a
37	299	2.8	1183	1 CNA_STAU	O53654 staphylococ
38	298.5	2.8	1296	1 ASAL_ENTPA	P17953 enterococcu
39	295.5	2.8	1256	1 MRP_STRSU	P32653 streptococ
40	293	2.7	1228	1 SLAP_BACST	P35825 bacillus st
41	293	2.7	1754	1 PMPB_CHLTR	O84418 chlamydia t
42	291	2.7	1258	1 ICEN_ERHME	P16239 erwinia her
43	290.5	2.7	1276	1 MRSP_CHLPN	O92899 chlamydia p
44	290.5	2.7	1637	1 MRSP_STAU	P80544 staphylococ
45	290.5	2.7	3562	1 PGCV_CHICK	O90953 gallus gall

## ALIGNMENTS

RESULT 1	ID	STANDARD:	PRT: 2249 AA.
AC	OMPA_RICRI		
AC	P15921		
DT	01-APR-1990 (rel. 14, Created)		
DT	01-APR-1990 (rel. 14, Last sequence update)		
DT	16-OCT-2001 (rel. 40, Last annotation update)		
DE	Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rompa) (romp A).		
GN	OMPA..		
OS	Rickettsia rickettsii.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;		
OC	Rickettsiaceae; Rickettsiense; Rickettsia.		
OX	NCBI_TaxID=783;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=N.		
RC	MEDLINE=90354033; PubMed=2117568;		
RA	Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;		
RT	"A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";		
RL	Infect. Immun. 58:2760-2769(1990)		
CC	- FUNCTION: ELICITS PROTECTIVE IMMUNITY.		
CC	- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A		
CC	- LAYER WITH HEXAGONAL SYMMETRY.		
CC	- P-TM: GLYCOSYLATED (PROBABLE).		
CC	- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	EMBL: M31227; AAA26380.1; -		
DR	PIR: A41477; A41477.		
DR	InterPro: IPR003858; rompa.rompb.		
DR	Pfam: PF02708; rompa.rompb.1.		
KW	Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.		
FT	SIGNAL	1	28
FT	CHAIN	29	2249
FT	DOMAIN	212	1180
FT	REPEAT	212	286
FT	REPEAT	287	358
FT	REPEAT	359	430
FT	REPEAT	431	505
FT	REPEAT	506	577
FT	REPEAT	578	652
FT	REPEAT	653	724
FT	REPEAT	725	799
FT	REPEAT	800	874
FT	REPEAT	875	949
FT	REPEAT	950	1021
FT	REPEAT	1022	1093
FT	REPEAT		

REPEAT 1094 1165 M (TYPE II).  
 REPEAT 1166 1180 TYPE I (INCOMPLETE).  
 SEQUENCE 2249 AA: 224333 MW: A9D6646C089DE087 CRC64:

Query Match 5.9%; Score 630.5; DB 1; Length 2249;  
 Best Local Similarity 23.2%; Pred. No. 1.9e-16;  
 Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

180 LADTMDLQYRTHAG-----HASTAV-----GAMSYAG-----HFSNAFGR- 218  
 1 MANISPKLFKAIOGLKALFTSTALIMLSSGALGAVGATATNNAAFSNVNGNN 60  
 219 -----ATAEGT-----YSLAVGLTATAKASSIVGSAOIGFAATAVGSGT 261  
 61 WEITRAAGVANTPAGGPONNMAFTYGGDYVTADAAIRIKAIN-----VAGTT 110  
 262 QVNLN-----RG-----IALGFSOVLKNDVNAANRAVAPDDNPIIDNRK 305  
 111 PVLGITONTVVGSITFTGNLLPVLNAGKSLTLGNNAVANAGFADADNY----- 163  
 306 ATRKAGADVPSIGNSNDISIRKIIIVGASADTDVANAOLKEAVRLANROTFKGD 365  
 164 -----TGLGIALGANAAL-----IIOAAPSKITLAGNID-----GGITTYKTD 205  
 366 DSNNEVEKGLKRTITTGGAQTSALTIDHNIGVONGDLKQOLAE-----TLTSLKVT-- 420  
 206 AAINCTIGNTALATVNVGAGATATLG-----GAVIKATTTKLTNAAVLTLLNANAVLGA 261  
 421 -ENLTANEKVY---GKRLTLDKIGFTNDMNGIDESKPYLDKDTGTHAGGOKITKLT 475  
 262 IDNTTGGNVGLNLNGALSOVTDIGNTSLATISVG-----AGATLGGAVIKATT 314  
 476 GVDDDAATPYGLKKVNOT-----AESALOPTVAKVKNKDADSDSIIVY---G 523  
 315 TKLTPAASVKEFTNPVVYGAIDNTGNANNGIVFTGNSV---VTGNGVFTNLATVNNCAG 373  
 534 KNNKPDG--TOVNTLKLKENGVDYTTETNGVTYEGT-----NONNGLT--VGNSTLNDG 575  
 374 LLQVGGVAVKANTINL--TDNASAVTFPNPVYVGAIDNTGNANNGIVFTGNSVTDG- 430  
 576 LSVKRTNSNKQIOVG-----ADGIFTDISNKPAGAGIENTRI 614  
 431 --IGTNTLATVNVGAGATLGGAVIKATTTKLTNAAVLTLLNANAVLGA-IDNT-- 485  
 615 TRDGIGFANNTSGLDANKPRLEPTINGAGKELTNVOSAINPATNGGOLDPMNRSTANT 674  
 486 GGDNVGVNLNGALS-----QVTGNIGNTSLATISVAGATATLGGAV-----IKATTT 534  
 675 EKSGSAAITKIDLYNLSOYPLTFAGDTGPNVTKKLGELLKVGKGTTPADLTKN---IG 730  
 535 KLTDAASAK--FTNPVVYVGAIDNTGNAN--NGIVFTGNSVTDGIDNTGNTSLATIS 588  
 731 VVADSTDSLVYKLAKTLDSDAVNTKLTFTASDKV---TVDS--GNNTATLQ--NGDLTF 783  
 589 VGAGTATLGGAVIKATTTKLTNAAVLTLLNANAVLGAIDNTGNDVNVGLNLNGALS- 647  
 784 SKONTGAPATPNS-KTIGVNDGLKFTDNNNGIALDGTYYTKKVGKAGKODGSLDKSKPYLD 842  
 648 --QVTGIDGNTSLATISV--GAGTATLGGAVIKATTTKITNAVASVK-----FT 693  
 843 KDKLVGEVEITNGIN-----AGKAITG--LSNLTLDAT--NATGHAUTOGLIV---DST 892  
 694 NPVVYVGTGIDSTGNANNGIVFTGNSVTDGIDNTGNTSLATISVAGATATLGGAVIKATTT 753  
 893 DKTRASTGIDVNLNAGFNKKNNGDAKDFSTVDYFINGNATFTAKVTED--OKASVADV 951  
 754 KLTNAASVLTLLNANAVLFG---AIDNTTGGDNVGLNLNGALSQVTDGIDNTGNTSLA-TI 809  
 952 NVDTGTTIHLTGADGNKNOIGVKTTLTLDKAGDKAINEFVSNGSGDKADLADNMLN 1011  
 810 SVGAGTATLGGAV-----VIKATTTKLT-----NAAVLTLLTNANAV----- 845

1012 TLAGEIRNKGADTALQTPQVKKVRENG--DDNDADTITVCKDAKT-----NOVNTLK 1064  
 846 -LGVAVDNTTGGDNVGLNLNGALSQVTDGIDNTGNTSLATISVAGATATLGGAVIKATTTK 904  
 1065 LKGRNGIDTQTNKDGVTYFTGINTQSLKAGNNT--TLNNGGL-----SKNTAGNQIOY 1117  
 905 LTNAAVLTLLTNANAVLTGAIIDNTTG---GDNVGLNLNGALSQVTDGIDNTGNTSLATISV 961  
 1118 GA-----DGVKFAKVNNGVYVAG--IDGTRITRDEIGFAGTN---GS 1155  
 962 GAGTATLGGAVIKATTTKLTDAAASAKFTNPVVYVGTGIDNTGNAANGIYFTGNSVTDGN 1021  
 1156 LDKRNPRLSDKGINAGKRTITNIOSEIAONSDAVYTGKTYDLETKLENKISSTAKTAQ 1215  
 1022 VGNNTA-LATVNVGAG---LLOYGVGVKANTINLND-----NASAVFTT 1062  
 1216 RSLHEFVADGQGNFTVSNPSSYDTSKSDVITTAGENGIT-----TKVAKGV 1266  
 1063 NPVVYVGTGAIIDNTGN-----ANNIGVFTGNSVTVGNVNTNALATVNVG-- 1106  
 1267 RVGIDOTKG-----LTPKLTVG-----NNGKGIYIDSONGQNTIT 1303  
 1107 -AGLLQVGGVAVKANTINLTDNASAVTFPNPVYVGTGAIIDNTGNANNGIY--FTGNSVTV 1163  
 1304 G-LSNTLANVNDKGSVYRTTEQGIIRKEDKTRASIVD-----VLSAGFNLQNGEAVD 1357  
 1164 GDIGNTALATVNVGAGITLQAGSL-----AANNIDGASRTLEFNGPLDGGKALP 1216  
 1358 F-----VSTYDTPAFQGNATTAATVYDDPDSKTSKYV 1390  
 1217 YFEKGAIANGNMILNVTNKLTAHLTIGTVAEINIGAGNLETIDASVD-----VT 1269  
 1391 DVANVDPTIEVKKKLGKVTITLTSTGANKFALSNOATGALVRAISIVD----- 1442  
 1270 ILNAQNIINFARBSVLYLSNL--TGCVYNNILL-----ADLVAPGADGTV 1314  
 1443 ---HLNPLSGDIQTAKASQANSAGYVDADGNK---VIYDS---TDNKYYQAKNDGT 1491  
 1315 VFNGGVGNL-----VGSNVAGTARNIGDGGNKNFTLLIYNAVITFDVNLGEGIONYL 1368  
 1492 VDKTKVEYAK-----DKLVAQAPDGLT--AQNNVKSIVNKEGVNDA 1531  
 1369 INKNADFTSTAFNAGAIQINDATYITIDANNGLNIPAGIOPRAHADQVLVLONSGGR 1428  
 1532 NKKGINEDNAFYVKEKASDNKTKNAATVYGLDLNAV--AOPLTFRAGDTGT--AKRL 1587  
 1429 TITIGANID-----PDMDDEGIYV--LMSVYAGKRLTIAG--GKTFGGAHKL 1471  
 1588 GEPYLTIG--GOTDNTKLTIDNNGIVAGTDFYVYKLAKDLNLNSVANAAGTKIDDKVSEV 1646  
 1472 -QTLTFRAGADGSP-----AGTFNTNIVLIDIT----- 1499  
 1647 DSSGOAKANTPVLSANGLDGKGVISNVGKGTDTDAANVQOLENVRNLGLGNAG---N 1703  
 1500 ---GQ-----LEIG-----ATTANVLENDVAOLTOGNCIGFID 1531  
 1704 DNAD-----GNOVINADIKDPNCGSSSNRTVYIKAGTVYLGKGNDEKLA---TGVO 1754  
 1532 ENAKNGAVTLLNNVNVAGAVO--NTGGTNGTLI---VLGASNLNRVANGIAMLKAGACN 1585  
 1755 YGVUDKODNAN--GDLSNVWVYVTDQDGSKKALLATYNAAGOTNITLNNPAIRDIRINGGR 1813  
 1586 VTIKAGKVRKIGEGGTNT-----LTLPAHFNLG-----SINTGGQALK 1628  
 1814 FFHYNDGNQPPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGROTQAGNQSIAID 1873  
 1629 LNFNNGSVSGV-----GTAANSVDITTAGATSPASSV 1663  
 1874 NQAT-----GDOSTA---IGTGNVYAGKHSAGIADPSTYKADNSYSVGNNOFTDATQTD 1926  
 1664 NAKGTATLGGTSTFANFTNTGAVTLAK-----GSITSFAKNVTAT 1704  
 1927 VFGVGNNTIYTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTGA---TGTVAG 1981

DB 1705 SF-VANSATIPNSLSLAFNSN--ITGC-----GTLITGANOYTTGT-GS 1746

QY 1982 FAGOTAVGAVSVC--SGAERRIONVAGEVSATSDAVNSOLYKATOGIANTN 2035

DB 1747 FTDTLTMTFTDGAAGAKSGCNLTIRSGSLDLSGVSTLAL-----VLTATN 1791

RESULT 2

OMPA-RICCN STANDARD: PRT: 2021 AA.

ID OMPA-RICCN

AC Q52657: P95591; P95593; P95594; Q52667; Q52668; Q52669;

AC Q52670: Q52674;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rompA) (rompA).

GN OMPA OR RC1273.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiidae; Rickettsia.

OX NCBI\_TaxID=781;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;

RA MEDLINE=94171067; PubMed=8125327;

RA Croquet-Valdes P.A., Weiss K., Walker D.H.;

RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia conorii (Malish 7 strain).";

RT Gene 140:115-119(1994).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=Malish 7;

RA MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Renseto-Audiffren P., Fournier P.-E., Barbe V., Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., Raoult D.;

RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.," Science 293:2093-2098(2001).

RL [3]

RN SEQUENCE OF 8-204 FROM N.A.

RP STRAIN=Indian tick typhus, MI, Malish 7, and Moroccan;

RC MEDLINE=97015921; PubMed=8862558;

RA Roux V., Fournier P.E., Raoult D.;

RT "Differentiation of spotted fever group rickettsiae by sequencing and analysis of restriction fragment length polymorphism of PCR-amplified DNA of the gene encoding the protein rompA.,"

RT J. Clin. Microbiol. 34:2058-2065(1996).

RN [4]

RP SEQUENCE OF 953-2012 FROM N.A.

RC STRAIN=Indian tick typhus, MI, Malish 7, and Moroccan;

RA Raoult D., Fournier P.E., Roux V.;

RT "Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rompA.,"

RT Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RL [5]

RN SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -1- PTM: GLYCOSYLATED (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/ROMP FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: 001028; AAA17405.1;

DR EMBL: AEO08674; AA103811.1;

DR EMBL: UA3794; AAB49549.1;

DR EMBL: UA3798; AAB49550.1;

DR EMBL: UA3806; AAB49551.1;

DR EMBL: UA5244; AAB49566.1;

DR EMBL: UA6918; AAB49663.1;

DR EMBL: UA8340; AAC35176.1;

DR EMBL: UA8344; AAC35179.1;

DR EMBL: UA8348; AAC35184.1;

DR EMBL: UA8345; AAC35189.1;

DR InterPro: IPR003858; romPA\_romPB.

DR Pfam: PF02708; romPA\_romPB, 1.

KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein; Complete proteome.

KW SIGNAL

FT CHAIN 1 38

FT DOMAIN 39 2021

FT DOMAIN 238 946

FT DOMAIN 1424 1528

FT VARIANT 60 60

FT VARIANT 76 76

FT VARIANT 86 137

FT VARIANT 126 133

FT VARIANT 952 954

FT VARIANT 1245 1245

FT VARIANT 1308 1308

FT VARIANT 1877 1877

FT CONFLICT 10 10

FT CONFLICT 92 92

FT CONFLICT 126 126

FT CONFLICT 137 137

FT CONFLICT 157 157

FT CONFLICT 368 368

FT CONFLICT 374 388

FT CONFLICT 640 640

FT CONFLICT 669 669

FT CONFLICT 793 793

FT CONFLICT 803 804

FT CONFLICT 809 823

FT CONFLICT 898 898

FT CONFLICT 908 908

FT CONFLICT 985 985

FT CONFLICT 1009 1009

FT CONFLICT 1013 1013

FT CONFLICT 1182 1182

FT CONFLICT 1314 1314

FT CONFLICT 1451 1451

FT CONFLICT 1624 1624

FT CONFLICT 1628 1628

FT CONFLICT 1872 1872

FT CONFLICT 1875 1875

FT CONFLICT 1878 1879

FT CONFLICT 1936 1936

FT CONFLICT 1936 1970

FT CONFLICT 1997 1997

SEQUENCE 2021 AA; 203328 MW; 327FCA2D7CB24668 CRC64;

Query Match 5.1%; Score 550.5; DB 1; Length 2021;

Best local similarity 21.5%; Pred. No. 1.6e-13;

Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;

QY 14 GEFMAVAEYAKSHSTGGSCATGCGSVFTLSFARIALAVLVGA-----T 60

DB 88 GYTTTADVADHITTAIVADTTPIG-----LTIAMONTVGSIVGNNLPTITAGKSLT 143

QY 61 LNSAVYAGIGSEADGKGGANARGDKSIATGIDIA-----QALSGSQTATIDNKIVHNSN 115

DB 144 LMG-----NNADAANHGFAPADNTGLCNIALGAMNALLIIOSPAAPKITLAGIN 195

QY 116 NNANIGAKASGHESTAIIG-DVLASGHASIAIGSDTLKREYQOISELLPIIRGOKAL 174

DB 176 GGIITVKTDAINIGTIGNTALATVNVGAGIATATLEGAITKATTTKLTMAASVL----- 249

QY 175 NDYQADNLOKRYRTHAQAAGHASTAVGAMSAKGFHSNAGFTR---ATAEGTSLAVG 230  
Db 250 -----TLIVNAVLTGADNTTGVNDVGVNLINGALISOVTG 285  
QY 231 LPTAKAASSIIVAGSMAOIGAFAATAVGSGTOVNLNRGIALGFGSOVLQK---DNDAVAA 287  
Db 286 NIGNTNALATISVAGAKATLIGA---VIKATTTKLTDNASAVTFNPPVVTGADINTGNA- 342  
QY 288 NVRATAPDNDPIDNRYKATRNKATDVFSIGNSGNDISIRKILINGAGSAD----- 340  
Db 343 -----NNGIVTFGTGSTVGNIGNTNA-----LATISVAGAKATLIGALIK 383  
QY 341 -----TDVAVNAOLKEAVRLA-----NROIFKGDSNN-----RVEKG 374  
Db 384 ATTKLTLDNASAVTFNPPVVTGADINTGNANNGIVTFGTGSTVGNIGNTNALATISVG 443  
QY 375 LKLTLLTGG---AOTSALTDHINIGVONGDLKVOLETLISLKMVT---TENITAN 426  
Db 444 AGKA-TLGGALIKATTTKLTDNASAV-----TFNPPVVTGADINTGNAN-N 488  
QY 427 EKVYTKRLTLDKIGFNDMGIDESKPYLDKDTGHAQ---GQTKTLTACVDDA 482  
Db 489 GIVTFGTGSTVGNIGNTNALATI-----SVGAKATLIGALIKATTTKLTDNA 537  
QY 483 ATYGOLKRYNQT-----AESALQFT-VKVDKNGNDANDSKITIVGRNNRP-DGTQ 532  
Db 538 SAVTFNPPVVTGADINTGNANNGIVTFGTGSTVGNIGNTNALATISVAGAKATLIGAI 597  
QY 533 VNTLKGKENVYDTTEINGVTFGL-----NONNGLT-VGNSTLNDGSLVAKNTSNK 585  
Db 588 IKATTTKLTDNASAVTFNPPVVTGADINTGNANNGIVTFGTGSTVGNIGNTNALATI 654  
QY 586 QIOVGA-----DG-----ITFTDISNSKPGAGIENTRITRDGIFANN 624  
Db 655 TYNVGAGATLEGAVIKATTTKLTDNASAVLILTNNAVLTGA-IDNTTGV---DNVGVNL 711  
QY 635 TGSID-----ANKPRLPTGINAGSKELTNVOSAI---NPATNGOL 663  
Db 712 NGALSQVTGNIGNTNALATISVAGAKATLIGAVIKATTTKLTDNASAVTFNPPVVTGAI 771  
QY 664 DEMANLSTANTKESGSAITIDLYLSOVLPLFAGDTPNTTKLGE-----I 711  
Db 772 D-----NNGNA-NNGIA-----TFGTGS---TVTGNIGNTNALATVNVAGL 809  
QY 712 LKVGSKTTADDLTKNNIGVAVSDTNSLJFKLAKTISDLDAVNTKLTASDKYTVDSG- 770  
Db 810 LRVQG-----GVVKSNTIN-----LTD---NASATTFNPPVVTGAI 843  
QY 771 NNTAKLQNGDLTFESKQK---TGATPATNS-KTIVGDLKFTDNGIALDGTITYTKDKYGF 827  
Db 844 DNTGNANNGIYTFGTGSTVGNIGNTNALATISVAGAKATLIGALIKATTTKLTDNA--- 900  
QY 828 AKQDSILSKRKYLDKDLKLVGEVEITINGNAGKATISNLTLDATNATNTAHVQGL 887  
Db 901 -----SAVTFNPPV-----VVTGADINTGNANNGIV-----TFGTGSTV 936  
QY 888 IYDSTDKTAAISIGDVLNAGENLKNNG---DAK---DEVSTYDVTDEFINGNATTAKYVDGK 943  
Db 937 NIGNTNALATVNVG---AGVTLQAGSLDANNIDF-CARSTLEF-NG-----PLDGG 983  
QY 944 ASKAVYDVAVNDGTTHLTG---ADGNKNOIGVKTTLTKTDAKDAKALFVSNSGD----- 996  
Db 984 GNAIY-----YFGAIALNGNALLNVTKLTPYHLLTGTVAEINIGAGNLEFAID 1034  
QY 997 -----DKALINKADIADNNTLAGEIRNTKGTADALQFOVKKREKENDDON---DADTI 1049  
Db 1035 ASAGVYTLNMADI---HRRAL-----DSALVLSNLTGCVN---NILLADLV 1077  
QY 1050 TVGKDAKTNOVNTLKLKKNGLDIQNTKGTVTFGINTQSILKAGN---NTTLNNGLSIK 1107  
Db 1078 APGVDEGTVEFD---GGVNGLNGSNAVAGAA-----RNIGDVGKRNKFNLLITNAVYIT 1128

QY 1108 NTAGNEQIOVAGDVKRPAKVNNGVAGIDGTRITRDEIGFAGTNGSLDKSPHLKSDG 1167  
Db 1129 D-----DVNLEIGQVNLNN---NADFTSSATPAAGTQIQUIDATYITDANNGNIN--- 1175  
QY 1168 IMAGKRTNIGSGEIAONS---ND-AVTGKRYDLTELENKISSTAKTAONSLEFSA 1224  
Db 1176 IPAGNIKFAHADQOLILQNSGNDRTLLIGANIDPNDDEGIVILNSVTA----- 1225  
QY 1225 DEGNNTVAVNPPSSYDTSTSVITTFAGENGITTVKNGCVAVNGIDQTKGLTPRLTIG 1284  
Db 1226 ---GKLTLLIGKTFGGANHLQDIY-FKGE-----GDFGATGTF----- 1261  
QY 1285 NNNKGIIVDSONGONTITG---LSMTLAVNTDKGSVTFTEOGKILIKEDKTRAASIYD 1341  
Db 1262 ---NTTNIVLD-----ITQLELGATLANVLEKAVOLTOVG----- 1296  
QY 1342 VLSAGFVLQNGEAVDFVSTYDVTVPADGNATTAKYTYDDTSTKSVYVDVAVNDTTEY 1401  
Db 1297 ---NIGG-----FLDFNAKNGTYTLNN-----NVNVAGT--- 1322  
QY 1402 KDKKLGVKTTTLTSTGT-----GANKFALSQATGDALVVASDIAVHLNLTSGDIOTAKGA 1457  
Db 1323 ---VKNTGTNNNGTLLVGLASNL---NRVNGIAMLKVG-----AGVNTIAGK- 1363  
QY 1458 SOANSSAGYVDADGNKRYIDSTDNKRYQAQKNDGTVDKTEVAKDKLVQAQTPDGTLAQ 1517  
Db 1364 ---GNVKIGEIGQGTNLT---TLPAHF---KLTSINKT-----GGQALKL 1401  
QY 1518 NVKSVINKEQVNDANKKOGINEDNAFVKGLEKASDNKRTKMAVYTGDLNVAQTPLTFA 1577  
Db 1402 N-----FMNG-----GSVSGVGTANSV 1420  
QY 1578 GDGTG-ARKLIGETLTKG---GQTDJNKLDNNIGVAVGTGTFVKLAKDL----- 1626  
Db 1421 GDITTAATSPASSVANKGATIGGTTSPAHFTTNNGAVTLAKGSTITSAKNTATSAFA 1480  
QY 1627 NLSVNAQ---GTXIDKGV-----SEVDS-----SGQAKANTPYL 1659  
Db 1481 NSATINFGNSLAFNSNTTSGTLLTGLANQVYTGTSFDTLLTLPEDGAKSGGNIL 1540  
QY 1660 SANG---LDGKGVISNVKKTDTDAANVOQLNEVINLGLGAGNDMDGQVNTADIL 1717  
Db 1541 IKSSTLIDLSG---VSNALVYVAT---NFDNM----- 1567  
QY 1718 KDPNSSSSNRVYIKAGTYVGKRG-----NND-----TEKLANG 1751  
Db 1568 ---NISPDTKYVISAETAGLKPTEKENVKITINDNRPVDFTPDASTLTLFAEDIAAG 1624  
QY 1752 GVQGVVDKQGNANGDLSNV-----WVTKQDKSGK-----ALLATYNAAGOT 1793  
Db 1625 ---VIDEDFAPGPPLANIPNANIKRSLELMEDAPGSDARQAFNFGMLPLOGADAT 1680  
QY 1794 NYLNN---PAEAIIDRINEGI-----PEFHVNDGQEPVVOGRNGISSAS 1837  
Db 1681 THLMQDVYKRSPTTAAVNNQVYVANSISSNTTALNAMAQKVOAGNKPVSGBEDMD- 1736  
QY 1838 GKHSVALIGFOAKADGEAAVAIGROTAQAGNOSIAGNQAOTG-----DOSIAGTG 1888  
Db 1737 ---AKFGAMISPPEVGNATQKMCNSIS-GYKSPDTGGTIGFDGFSVDLVLGLA 1785  
QY 1889 NVYAGK---HSGAIGDPSYKAD-----NSYSVGNNNOFTDA----- 1922  
Db 1786 YTRADDIDIKLNKKTGDKNKNVESNIISLYGLASVPEYLEVEFIASYSPDKTRSSRYI 1845  
QY 1923 ---TOEDFVGNNITVTESSVAL-----GSNSAISAGTHAGQAKKSDGTAGT 1969  
Db 1846 ATTLFVGYGTANGKYSSEYTGQOLMAGYTYMSEMINILPLAGLAKYSTIKKSYKECT 1905  
QY 1970 TTTAGATGYKFGAGOTAVGAVSVGASGERRIQNVAAGEVASTSDAVNGSOLIKATO- 2028  
Db 1906 TY---QNLTVKKNYNTFPDLGAKVS-----SNINVEIYVLT-----PELYAMVDY 1949  
QY 2029 GIANATNEIDHRIHONENKRNANNGISSAMAMASMPQA 2064



Db 1950 AFKNVSAIDARLO-----GMTAPLPTNSFKOS 1977

## RESULT 3

YDBA\_ECOLI STANDARD; PRT; 2003 AA.  
 AC P31666; P76087; P76856; P76857; P76859;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ydbA.  
 GN YDBA OR B1401/B1405.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia coli.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-9742617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97251357; PubMed-9097039;  
 RA Alpha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [3]  
 RP SEQUENCE OF 464-2003 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-92190338; PubMed-1665988;  
 RA Moszer I., Glaser P., Danchin A.;  
 RT "Multiple IS insertion sequences near the replication terminus in  
 Escherichia coli K-12.";  
 RL Biochimie 73:1361-1374(1991).  
 CC -1- SIMILARITY: TO S. TYPHIMURUM ORF NEAR CYS6 (NC P25928).  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. THE GENE CODING FOR  
 THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
 BETWEEN AMINO ACIDS 839 AND 840.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE000237; AAC74483.1; ALT\_SEQ.  
 DR EMBL; AE000237; AAC74487.1; ALT\_SEQ.  
 DR EMBL; D90778; BAI1509.1; ALT\_SEQ.  
 DR EMBL; D90778; BAI1880.1; ALT\_SEQ.  
 DR EMBL; D90779; BAI1881.1; ALT\_SEQ.  
 DR EMBL; X62680; NOT\_ANNOTATED\_CDS.  
 DR Ecogene; EC11307; ydbA.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 489 I -> V (IN REF. 2).  
 FT CONFLICT 495 I -> V (IN REF. 2).  
 SQ SEQUENCE 2003 AA; 205949 MW; B83A12CB53220EE CRC64;

Query Match : 5.1%; Score 544; DB 1; Length 2003;  
 Best local similarity : 21.3%; Pred No. 2, 7e-13;  
 Matches 415; Conservative 242; Mismatches 695; Indels 600; Gaps 102;

QY 274 GSOVLQKNDVNAANVRAAPDDNDPDKRYKA-----TFKNGATDVFSIGNSNGD 325  
 Db 122 GDEIIPDDPDTPTPKPVAFNNVDVLDKTEKTLTRDSVFTTENADGTSISLDSNGRK 181  
 QY 326 SIKRKIIINGAGSADIDAVVAQLKAVRLANRQITFKGDDSNRRVEKGLGTLITGGA 385  
 Db 182 A-----TINLMQIDE-----ANNVTALEG-----VSADG 205  
 QY 386 QTSALDHNIGVONGPDKGLVQLAEFLTLKMTTFENLTAKEVTVGKTRLPD-KIGFT 444  
 Db 206 ATKMQIHNELVITGD-----NATVNNN--GKT--TVDGKDSIG 241  
 QY 445 NDNGIDESKPYLDKPDGTHAGGOKITKLTAGVDDAATYGLKKNQTAESALQTFV 504  
 Db 242 TEING-NGKVIODGDLVSGGCHGID-----ITGDSATVDN-----KGTMTV 283  
 QY 505 KKYVKNQNDANDSKITIVGKNNKPDGTQVNTLKLGENGVDTTEGTGVLNONGL 564  
 Db 284 TDESMEIQIDGKALIV--NNEGESTITNGCTGTQINGDAAANNNGKTTVDGKDSGT 340  
 QY 565 TVGNSTLINNGLSVKNKNSKQIYGVGADGITFTDISNKPAGIENTFTLRIGIGFANN 624  
 Db 341 ET-----NKNNGKVIODG-----DLVSGGCHGID-----ITGDS-ATVDN 375  
 QY 625 TGSIDANKPRLPPTGINAGKELTNVOSAIPNTPANGOLDPNNRLSTANTERSGAATIK 684  
 Db 376 KGMITVDP--SIGIYQVDSDAVNVNNEGSAITNG-----TGTQINGDAAIAN 423  
 QY 685 DLYNLSOVLTPFAGDPNATKRLGELLKYGKGTATADLTNNIGVADSDNSLTVKL 744  
 Db 424 ---NNGKTVYDGDSTGTEIAGNNKRYIQ-----DGDLVSGGCHGIDIGDSATV-- 471  
 QY 745 AKTLSDDAVNTTILASDKVTVYDSGNNTAKLON--GDLTFPSKONTGA-----TPATNSK 797  
 Db 472 -----DNKGMTVTPDESIGIYQVDSDAVNVNNEGSAITNGCTGTQINGDAAIANNSG 524  
 QY 798 TIGVGLKFTDNNNGIADGTYITKDKVGFAGKODGSLDSKPYLDKDKLVGEVEITNG 857  
 Db 525 KTVYDG--KDSGTCTIAG-----NIGIVNIDGSL-----IVTGG 556  
 QY 858 INAGKAITGLSLTLDATNATGHTVGLGIVDSTDKTRASIGDVLA--GFNLKNNGDA 916  
 Db 557 AH-----GVENIGNGCTVNNKGDV--VVSdT-----GSLGVLINBGAATVSGTGV 600  
 QY 917 KDRVSTIYDVDFINGNATTAKTATYVDGKASVAVDVANDGTTILTGAD--GKNQIGVKT 974  
 Db 601 N-----VSNEATGFSITNGSKVSLAGSMOVGFES--TGVDINGNNNSV-- 642  
 QY 975 TPLTKTDK--GPKAIFNSVSGDDKALINAKDIAONLMTLAGEINNTGTADTLAQTFQ 1032  
 Db 643 -TLAARDLAKVGGAKAIGIN--SGD-----ANTVITGVAVLDKKTADNAEYTF 690  
 QY 1033 VKVYKEN--GDDNDADDTIVGDAKTNOVNTLKLKGKNGLDIOTNK-DGVTFYFGINTOS 1089  
 Db 691 DPSVGIWVSDNN-----VTLDGKLTVVSDSEVSR-----QSNLFDGSA-----EKTS 735  
 QY 1090 GLKA--GNNTLNNNG-----LSIKNTAGNEIOVGADGVAFKAYNNG-----VYAGG-- 1135  
 Db 736 GLAVVIGDGNVTNNNGGELLEGERNAL-----ADGSQVTSILRTSYSTSVYVSGESS 787  
 QY 1136 --IDGTRITRD--EIGFAGTNGSLDSKPHLSKDGINAGKRTTNIGSEIQAONSDAVT 1192  
 Db 788 VYLNGDTTISGEPLPGAGV-----TRVDKALLEIGSG----- 821  
 QY 1193 GGRITDKLETKLEKISSTAKTAQNSLHFSVADEQNNFTVSNPYSYDTSKTSDV--IT 1250  
 Db 822 -----ATLTMQDIDSEFHG-----TRVEIIONG 846



Db 247 NATANALNLOAGGT-----INFNGIDGTGRVLILSKNCAATDFNVTSIGSNGNK-GI 299

QY 851 VEITTINGINAGKAITGLISLTLDATNATGHTVQLGIVSDFTKRAST-GDVLNAGFN 909

Db 300 IELMTVALINGOLINAGPANAIVIGTNN--GAGRAAGFVSVGDKAKATIDGQY----- 351

QY 910 LKNNGADPEVSTIDYDFINGNATKATYTDGKASKAYAVYVNDGTTILHTGADGKNQ 969

Db 352 ----AKDNY-----IOSANAN--GOVNF-----RHIVDVGIDGTAFKTA-- 386

QY 970 IGVKTTTLTKDADKDAKINF-SVNSGDCKALINAKDIADLNLTLAGIRTKGTADTAL 1028

Db 387 ----STVAITONS-----NEGTTDFGLAAOVYVPTMTLTGNTFGDANNPGMTA--GV 434

QY 1029 QTFQYKRVKRENGDDNDADTTVCGKADAKTNOVNTLKLKNGLDIDQTKDGTVEFGINTQ 1088

Db 435 ITFPAANGTLASASADANY-----AVTNMTITAIASGVGVVQLSGHTAELRLG--NAG 485

QY 1089 SGLKAGNNTTLANNGLSIRKTAGNEQIOVGADGVKFAKVVNNGVYAG--IDGTRITRD 1145

Db 486 SVFKLADGTIVNGK-----VNGTFLVG-----GVLAAGAITLDSATTTGD 526

QY 1146 EIGFAGTNGSDLSKSPHLSKDG--INAGKKITNIQSGEIAONSNDVATGKTIYDKTE 1202

Db 527 -IGNGGGGAAIDST--LANDATKTLTGAGANIISANGTI-----NFGANGTI----- 573

QY 1203 LENKISSTAKTAQNSL--HEFSVADGQNNETVSNPYSSYDTSKTSQVITFPAGEINGITT 1259

Db 574 ---KLTST---QNNIYVDCDLAIATDQTVGVDASS-LTNQOTLTIISGLIGANNNTL 625

QY 1260 KVNKGAVRVGIDQTK-----GLTTPKLTVGNNN-----GKG-IVDISQ 1296

Db 626 ---QOFNIGSSKTLNGNVAINELVGNNGSVQFAHNTYLTITFTNAAGKTIIEPV 681

QY 1297 NQONTITGLSNTLAVTNDKGSVRTTEOGKTIKDEKTRASTIVYLSAGFNLOGNGEAV 1356

Db 682 VNNNTTLAAGTILGSAANPLAINGSKG-----ARADVLAV-----GECV 723

QY 1357 DEVSSTYDFV-----NFDAGNATTAKYTYDDTSKTSKVYDVYVNDTTIEVKDKKL 1406

Db 724 NLYATNITTTDANVSFEVFNAGKNIIVSGTVGGQGNKFNVTALD---NGTIV---KFL 776

QY 1407 GYKTTTLTSTGANKFALSNATGDAKYA--SDIVAHNTLSGDIQAKASQANSQA 1464

Db 777 GNATFNGNTTIANSTLQISGNYTADFLASADGTGIVEFNTPINVLTKQAVPVNAKL 836

QY 1465 GY-VIADGKVIYDSTDNKYQAKN-DGTVDKTEKVAKDLIAQAQTPD-----GTIA 1515

Db 837 QITVSPGPNVYV-----NEIGNAGNYHGMATDTIAFENSLSGAVLELPSCIFPNDAQNTI 891

QY 1516 QMNVKSVINKEOVNDANKKQGINEDNAFYKLEKASDNKT---KNAAVTVG---DLNAV 1569

Db 892 PLTIKSTVGENETA-----EGFSVSPSVIAGVDSYIADQVIGDQNNIITGLISGNGII 945

QY 1570 AQTPLPFADDTGTAKKIGETTLTGCGQDPTKLDNNNIGVAGVDTGCF--TYVLAKDLTL 1628

Db 946 VNATTLTYAG-IGTINNQG--TYVLSGVPNTPG--TYVIGSGTIGASKFQVFTTIDYNNL 1002

QY 1629 NSVNAAGTICID--KGVSFVD-----SSGOAKANTPYLSANGL----- 1664

Db 1003 GNIAIANTTINDGVYTTGGIAGIAGIDPQKILGVSNGNANVRFPADGIFSNSTMI 1062

QY 1665 -----DIGKVISNVGKGTQDPAANYQOLNEVRNLIGL-GNANGNDNADGNOYV 1712

Db 1063 VTTKANNGTLYLGNAFVGI--GSDDTFVASV--RFTGSNNGAGLKGNTYSOIVTEGTVN 1119

QY 1713 IADIKKDPNNGSSSNRFTYKAG--TYVLGK-----GNN---DREKLA 1749

Db 1120 LGIV-----NSNVILGSGTTAINGRIDLLTTLTLPAGSTGWNNTSLETTTLTA 1169

QY 1750 TG-----GYOVGVDXDKG-----NANGDLSNVWYKTOKDQSKR----- 1781

Db 1170 NGNIGHIVIAEGQVAVATTTGTTINVDNANANFSGTGYTLTIGGARRNGTLGPNFT 1229

QY 1782 -----ALLATYNAG--QTNVLTNNPAPARIDRINQIGRFFH-VND 1819

Db 1230 VTGNSRFVNYGLIRANODVYITRTNNAENITYVNDITNSPFGAPGVGONVTFVATNTF 1289

QY 1820 GNOEPVVOGNGIDSSASGKSHVAIGFOAKADEAAVAIGROTQAGNOSTAIGDNAQATG 1879

Db 1290 AAYNNILLAKNSDASNANFVGTIYTDISAATTNNQDLVAKDIOQLGNRLGALK----- 1342

QY 1880 DQSTAIQGNVAVGKSHGAIQDPSTYKA-----DN-SYSGNNNOTTDATQIDVPEVGN 1933

Db 1343 ----YLTQPEMV--GSEAGAI--PAAVAGADEAVDNNVAGTWARPFYTDHQSCKGGL--- 1392

QY 1934 IYTESNVALGNSALSASQTHAGTQAKKSDGAGTTTGTGANGTYKGFAGQTAAGVASV 1993

Db 1393 -----AGYKAKTGYIYGLDTLAN-----NNLMIGR-AI 1420

QY 1994 GASGAERIRIONVAAEVSASTDAVNG--SOLYKATQIGIAN--ATMELDRHIOENK-- 2047

Db 1421 GITTDTLKHQDYKGD-----KTD-VNGFSFLVGAQGFENFPAGSAIFSLNOVKNSQ 1475

QY 2048 ----ANAGISSAAMAMASMPQATIPGRSMYTGCIAT-----HNGQAVAV-----GLSKLSD 2094

Db 1476 RYFEDDANGNSKQIAAGNYD-----NMTFGGNLTGYDYDNAMQGVLYTPMAGLSYLS 1528

QY 2095 NGQWVFKEG-----SADTQGHVAGAVGAG 2119

Db 1529 SDE-NYKETGTVANKOVNSKFSRDTLIVGAKYAGG 1564

RESULT 5

BIG\_A\_SALTY STANDARD; PRT; 1953 AA.

AC P25927; P25928; Q9XQ03;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Putative surface-exposed virulence protein bigA precursor.

GN BIGA OR STM3478.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI\_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SCS1412 / ATCC 700720;

RA MEDLINE=21534948; PubMed=11677609;

RA McEldelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal M., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

RL Nature 413:852-856(2001).

RN [3]

RP SEQUENCE OF 1-765 FROM N.A.

RC STRAIN=LT2;

RA MEDLINE=91100301; PubMed=1987123;

RA Wu J.Y., Siegel L.M., Kredich N.M.;

RA "High-level expression of Escherichia coli MADPH-sulfite reductase: requirement for a cloned cysC plasmid to overcome limiting streptococcal cofactor.";

RL J. Bacteriol. 173:325-333(1991).

CC -! CAUTION: Ref.3 sequence differs from that shown due to frameshifts in positions 414 and 732.

CC This SMS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-ch.ch](mailto:license@sib-ch.ch)).

DR	EMBL; AF133696; AAC89458.1;	-
DR	EMBL; AE008859; AAL22340.1;	-
DR	EMBL; M64606; AAA27042.1;	ALT_FRAME.
DR	EMBL; M64606; AAA27043.1;	ALT_FRAME.
DR	PIR; C39200; C39200.	
DR	PIR; D39200; D39200.	
DR	StyGene; SG10437; bigA.	
KW	Viruslike; Repeat; Signal; Complete proteome.	POTENTIAL.
FT	SIGNAL	1
FT	CHAIN	28
FT	DNAIN	1933
FT	DOMAIN	101
FT	REPEAT	101
FT	REPEAT	103
FT	REPEAT	104
FT	REPEAT	114
FT	REPEAT	123
FT	REPEAT	133
FT	REPEAT	134
FT	REPEAT	144
FT	REPEAT	145
FT	REPEAT	146
FT	REPEAT	155
FT	REPEAT	167
FT	REPEAT	177
FT	REPEAT	188
FT	REPEAT	189
FT	REPEAT	199
FT	REPEAT	200
FT	REPEAT	210
FT	REPEAT	211
FT	REPEAT	221
FT	REPEAT	222
FT	REPEAT	232
FT	REPEAT	233
FT	REPEAT	243
FT	REPEAT	252
FT	CONFLICT	207
FT	CONFLICT	207
FT	CONFLICT	514
FT	CONFLICT	514
FT	CONFLICT	1698
FT	CONFLICT	1698
FT	CONFLICT	1795
FT	CONFLICT	1798
FT	CONFLICT	1836
FT	CONFLICT	1837
SO	SEQUENCE	1933 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match	4.68; Score 493.5; DB 1; Length 1953;
Best Local Similarity	20.68; Pred. No. 2e-11;
Matches 431; Conservative 244; Mismatches 742; Indels 677; Gaps 100;	

```

QY 295 DDMQFDLNKFKTEFNKGATDVFESLGNNGNDISIRKLIINVGAGSADTVAVNAQLKEAVR 350
Db 244 DDDTFPPDSV-ITFENSGV-----IDKGDITLFDSEFKLDNGSVLEGA-----286
QY 355 LANRQTFEKGDDSNRRVKEGKLTLLTGAQNSA-----L 396
Db 287 -----WNSQDNOMQTLTADKTLNTGMDVIDANAAYIEGTQENGLYMKYDSRGYLI 341
QY 391 TDHNIIGVYONGD-----GLKVQLAETFLSLKAV-----TTENITJANEKVYTKRPLTDK 440
Db 342 ADQNTTIVISGDDQAHNSRQMDIS-GQDRKGVYISGDRIVNTLTGDSVSYGAGVWISG 400
QY 441 IGFYDMNGCIDESKEYLDKDTGHAGGOKITKLTJAG-----VDDDAATY-----485
Db 401 DGTWTWISG---HSTYDMMATGALISNGTTFADIDIAVSOGGTAIIIIDGNATIKNTG 456
QY 486 -----GOLKKNQTAESALOTFTVKYDKRNGNDANDSKIIITGVKRNKPDG 530
Db 457 TSDISGAGSTGVIDGNARVNNDG-----MITDGTGCGHIITDNNVIDNAGSTVSG 511
QY 531 TQVNTLKLKGEKGVADVTETNGVYTFG-----LNQNNGLTVGNSLTLNDGSL-----VKNT 581
Db 512 ADATALITEGNDL-VINEGQOTISGAVGTRIDGDADHTTNTGDIADVAGGSAAVIING 570
QY 582 NSMKQIQV-----GADGITFTDINSKRKGAGIETWTTITTDGIGGFANNTGSLDANKPR 634
. . . . .

```

D	b		571	DNGSLTQADLLVLTGAGCIIITYGNGEAKNTG--NATVRDASVGR-----	615
O	y		633	LPIPTINGMKGRELINVOASINPAINGGLDDPMNLSTANTIEKSSAATTIDLYNSOVL	694
D	b		616	---VVAGER-----NTEFNKKGDID-----VSLNCTGALVSQ--DMSQOVL	650
O	y		695	---TFPADTGPNVTYTKLIGELIKVGGKTADDLPTRKNICGVDSFDSNLS	740
D	b		651	PDDIMWVSQOSEGVSESSATSVS-----GDSNAVDIT-GNVMSIADYGODDL	698
O	y		741	TYKLAKTLSDDAVMTKTLTASDKVYYVDSGNNTAKLONGLTFESKONTGTAPATNSKTTG	800
D	b		699	---AAAPPLTG---VVVGNGNIVTIL-NGALNIDDDNLS---ATGQOYL	739
O	y		801	VGLKFET-DNNGIALDGTTYITFKKVGFAPKODGSLSKSKRYLDKDKLKVGEVITTINGIN	859
D	b		740	VVGLSVTGGDDNDVELDESGINITH-----SEDLDTGSADIT--GIS	778
O	y		860	AGGRAITGLSNTLTDAINATT-GHYTOLGIYDSTDKTAAISGD-VLANGFMLKNNGDA	916
D	b		779	VSGNSTVTLNGHSHITDTVVGGHV-----VLARVNGSGSLTLDGDSVDVAVNSIPTG--	832
O	y		917	KDFVSTYTPTVFINGMAFTAK-----VYDDKASKVAAD---VNVDGTHILTGDGKK	967
D	b		833	---YTTVALLMAGEEGSIENKCDITSHGYVIYRADNGSEVSNCGDILLVYATSNNSE	889
O	y		968	MOIGVKITTLRTDAKGDKAINEFSVNSGDDKALIN-----	1007
D	b		890	DR-----AAIRASGESAVHN--KAGGDITLISDOTPGSGSIEVYPLKWTHHTFYAM	941
O	y		1003	-AKDIADMLNLAGEIRRTKGTADTALOTPOVKVKEENG-----DDDNADDTTV	1055
D	b		942	HASDGDVAVND-EGATIHLOAGYGYTASRGKALMEGNIYIDLGPVLTDEDENITTSY	1000
O	y		1052	GKDAKTNQVNLTKJGKNKGIDICTNKOGVYFPFGINTOSGLKAGNNTTLNNGSISTKNTAG	1111
D	b		1061	W-----QPSSLYLTSSGGWA GSTIDAGDAT-AINT-----GNTTVNAGFGMALNG	1044
O	y		1112	NEQIOVGA-----DGVK-----FAKYNVGVAGIDGTRITRIDEIGFAGINGSID	1157
D	b		1047	GTAINOGETTLTADGYTGGADELVGMAALNGGVVINDISGVINIDAD---YGOAFSLD	1102
O	y		1158	KSRPHLSMDGINAGCKKITINLOSGETANONSNDNAVGTGKIYDLKTELKNSISSTATQAONS	1211
D	b		1103	SSSYTIINNCSINLNGSPM-----DDTBSHMGC-----PPTDKIWIOS	1136
O	y		1218	LHEFSVADDEGONNETVSNPYSSYSTKTSDVITE-AGENGITTKYKNGVVRVG-IDOTK	1274
D	b		1140	L-----PESGSDSPRTSPDTEFFTAG-----TLANYGETELNGVDVNG	1177
O	y		1275	GLTPKTLVGNNGCKGVIDSONCOMTITIGSLTIANLVINDKGSVRTTBQGKIINDEKXT	1334
D	b		1178	G-----WLINEMASLTV--NGIYVITINGANAALAN-----	1205
O	y		1335	RAASIVDVLISAGFMLONGEADEVFSYTYDV-NFADGNATTT---AKVY---DDTS	1383
D	b		1206	---XGTDLA-DAISPMWSLEFEADGSTITDOLLTLINGVYTYTYNNGDPTG	1249
O	y		1384	KTSKAYVD---VANVDITILEYKAKKL-----GVKITTLLASTGTGANKFALSNOATGD	1432
D	b		1250	SIACTSYQOEIVMTGMTVAEDKSILVSGSFYFNYNEEDATLITNSGSAVE---GGEENTII	1305
O	y		1433	ALVYASDIVAHMLNTLSDIDDTAKGASOAANSAGUYADGNKRIYOSTD-----	1480
D	b		1306	NLTPANDSLTOVN-SGITITATNGYSALTITVNGSNDP---KIMINTATGVINGINPDAPL	1360
O	y		1481	---NKIYOAKNDGTVDKTEVANDKLVAQAQTDPDGLLAOMNKVSYINKEOVN---DANK	1534
D	b		1361	INLRGYNFNEGOSTIN---YOGDNAVA---ISGCTSXY-VINLVNSGTINVGEGQKE	1411
O	y		1535	QGINEDNAF-VKLEKRAASDNKRKMAAYVYGLDLNVAQIPPLTFAG-----	1578
D	b		1412	DGINGTCIGICKNGAKTTINNTAD---GVINYADDSVAFGKTKAIINGEINILLC	1466

QY 1579 DTGTTAKKLGELITKGGGOTDNKLTNDNT--GVVAGTGTGVLAKD-----LTN-LN 1629  
 DB 1467 DSGCIVAPGTT-----GTQNHNGTADIVIPDATTAPTEGSIPTPPADPNAPQSLNIV 1522  
 QY 1630 SVNAGGTIDKGVSVFSSGQAKNTFVLSANGDLGGKYSANVGKTKT-----1681  
 DB 1523 GTNAG-----SSGTLKANNLIVIGDN-----VKDTGFTSTADITVYVDAF 1565  
 QY 1682 DAANYQOLNEVENLGLGNA-GNDNADGN-----OVNIADIKKPNSSS 1725  
 DB 1566 TGSNIQADNITSTSVVNMAGSQDADGVDTMTKNAYADVATDSVSDVAQALDAGYT 1625  
 QY 1726 SNR---TVIKAGVILGGKGNDEKATAGVQVGVKDGNGANDLSNVWVKQKDGSKAL 1783  
 DB 1626 NNELYTSILWGT-----TAEINLSALKOV-----SGAQTATVFEARVLSNRTM 1669  
 QY 1784 LATVYAGGTNTLTNNPAEADIRINEGIRFPHVNDGNOEPVQGRNGIDSSASGHSVA 1843  
 DB 1670 LA--DAAP-----IKDG-----1680  
 QY 1844 IGFQAKADDEAFAVIGROT-----AGNOSIAI-----GDNAQATGDSI 1883  
 DB 1681 LAFNVYAKDPPAEIGNDTQYDMLAROTLDLTASQNLTEVGIARLDGDSKTAGDGL 1740  
 QY 1884 AIGTGVYAGKSGAIGDPTSKADNSVYGNNOFTDQTDVGVGNNTVTSNSVA 1943  
 DB 1741 TGGYSOFELKHSMAFDEGLANNLSLRVDVH-----LDSSRYAVAGDVAKIADSPMROOY 1796  
 QY 1944 LGSNSAISGTHAGTQAKKSDGTAGTTAGTGTGKAGAGTAVGAVSGASGAEIRIO 2003  
 DB 1797 L-----ERSEB-AKFTTMMGDALVTPYAGKFRHTMEDGKESAGDF 1840  
 QY 2004 NYAGEVSATSTDVANGSQL-YKATQG-IANATNELRHRIHONENKANAGISSAMAMSM 2061  
 DB 1841 NLSMNSGNFTAVDSITGLKLDYAGKDGWATATLEGCPMLYSKSGQRATSLQGAAG---- 1896  
 QY 2062 PQATYTRGSKMTGTGATNHNGQAVAGLSKLSDC-----QWFKNGSAD 2107  
 DB 1897 -OSFGVDGQKGGGV---NGLATIGV---KYSNPTALHLDAYOW--KEDGISD 1941

RESULT 6  
 OMPB\_RICPR STANDARD: PRT: 1643 AA.  
 ID OMPB\_RICPR 053020: 092CM0:  
 AC 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
 DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
 GN OMPB OR SPAP OR SPA OR RP704.  
 OS Rickettsia prowazekii.  
 OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;  
 OC Rickettsiaceae: Rickettsiae: Rickettsia.  
 OX NCBI\_Taxid=782;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RX MEDLINE=91043972; PubMed=2122457;  
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
 RA "Characterization of the gene encoding the protective paracrystalline-  
 RT surface-layer protein of Rickettsia prowazekii: presence of a  
 RT truncated identical homolog in Rickettsia typhi.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BREINL;  
 RA Moron C.G., Yu X.J., Walker D.H.;  
 RA "Sequence analysis of ompB of Rickettsia prowazekii.";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Minkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 RN [4]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RA "Mapping of monoclonal antibody binding sites on CNR fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii.";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [5]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RX MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent mutant  
 RT deficient in processing.";  
 RL Integ. Immun. 60:159-165(1992).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 DR EMBL: M37647; AAA26390.1; ALT\_INT.  
 DR EMBL: AF161079; AAD42234.1;  
 DR EMBL: AJ235273; CAAL5140.1;  
 DR InterPro: IPR003858; rOmpA\_rOmpB.  
 DR Pfam: PF02708; rOmpA\_rOmpB; 1.  
 KW Antigen; S-layer; Cell wall; Complete proteome.  
 FT CHAIN 1 1328  
 FT 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT VARIANT 1329 1643  
 FT VARIANT 257 257  
 FT VARIANT 1010 1010  
 FT VARIANT 1450 1450  
 FT VARIANT 178 179  
 FT CONFLICT 191 201  
 FT CONFLICT 212 212  
 FT CONFLICT 313 313  
 FT CONFLICT 1104 1104  
 FT CONFLICT 1123 1123  
 FT SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.5%; Score 480; DB 1; Length 1643;  
 Best local similarity 22.2%; Pred. No. 5.2e-11;  
 Matches 396; Conservative 222; Mismatches 619; Indels 544; Gaps 97;  
 QY 581 TNSKQIOVGADGTFITDISKPGATE-----NTTITTDGIGFANNGLDANKPRL 635  
 DB 18 TASTATVAGFSGVAM-----GAAMQYKRTTAAATTDGIGFDOAG--ANTPVA 66  
 QY 636 PTGTINAGKELTNQASINPAT-----NGOLDPMRLSTANTEKSSAATINDLY 687

Db 67 PMSVITA-----NANPEITNTNGHLSIFLDTFANDLAVTINEDT---TLGFTT 113  
 QY 688 NLSQVPLEFAGDGTGPNYTKLGEILIKVGGKTTAD-----LTKNNIGVADSTD 737  
 Db 114 NIAQAKFE-----NFTVAAGKILNITGGITVQESNNTIMQNALTRYHGGALIANAD 167  
 QY 738 ----NSLTVLAKTLSDLDANVRK-----TLASDKVTVDSGNNTAKLQNGDLTFESKON 787  
 Db 168 LSGISITFEAAPSVEFENLNPPTQEARPLTIGANSKI-VNGGNGTININTEFIQVS-DN 225  
 QY 788 TGAATPANSKTIQVYDGLKFTDNNIGALDGTYYITKKQVFGAKODGSLDKRPILDKKK 847  
 Db 226 TFA-----GIKTINID-----DCGLMFNSTP-----DAANTL-----NLQ 256  
 QY 848 VGEVEITTINGINAGKAITGLSN---TLTATNATGHTVQLGIVDSFDTKRAASI---G 901  
 Db 257 VGGNTINPENGIDGTGLVLSKNGAATEFNTVGLGSKLK--GIIELENTAAVAGKILSOG 314  
 QY 902 DVLNAGNKLNN-GDADEVSTYDTYDFINGNATTAKVYDGAKSVAYDVNDG--TTI 958  
 Db 315 GAANAIVGTGNGAGRAAGFTVSYD-----NGNAATISGQYAK-NMVIQSANAGQVTFE 368  
 QY 959 HLTGADGNKNOIGVYKTTTLTKTAKGDKAINFVSNGDDALINADIANLMTLAGETR 1018  
 Db 369 HIV-----DVLGIGTTNKT-----ADSKVITTENSNFGSTNF-----401  
 QY 1019 NTKGTADTALQTOVKKVEN--GDDDNADT-----ITVGRD---AKTNOVN 1061  
 Db 402 ---GNLDTQIVPDTKILKNFTIGDYKNNNGNTAGVITFMANGALVASIDPNTAIVN-IN 457  
 QY 1062 TTKLGR-----NGLDIOTNKGCTVTEGINTOSGLKAGNNTTLNNGLS 1105  
 Db 458 AIEEGAGVVELSGIHIAELRLNGSGSIFKLADGTIVNG-----PVONALMNNAL 510  
 QY 1106 IKNTAGNEQIOVAGVYKFAKVNNGVYAGIDGTTTRTRETREIGFAGNGSL-----PK 1158  
 Db 511 ---AGSIQ-----LDSALITGD-IONGGVNALQHTILANDA 544  
 QY 1159 SKPHLSKDG-----INAGKRITNIOGSELAONSNDVATGKHYDEKTEL 1203  
 Db 545 SK-ILALDGANIIGANVGAIHFQANGTIKLNTQ-----NNIVNFDL-DITTK 594  
 QY 1204 ENKISSAKTAONSLH-EFSVADQGNFTVSPYSYDTSKTSVYITRAGENGITTKVN 1262  
 Db 595 TGVVADSSLTNNQTLTNGSIGTVANTKTLAG--LNTGSKT---ILNAGDAIAMELVI 649  
 QY 1263 KGVAVRGIDQTKGLTTPKLTGVNNGKGIYIDSONQONTGLSNTLAVNTDKGSVRT 1322  
 Db 650 ENNGSVQOLNNTYLTIKTITMAAN--QGOTIYAADPLNTWT---TLADGTN-LGSA--- 698  
 QY 1323 ECKKIIEDEKTRASIVDLA--GFNLQNGEAVDFVSYD---TVNFADGNATTTAK 1376  
 Db 699 -EMPLSTIHFKATKANADSLILNKGKVNLYANN---ITTNDANVCSLHPSGGSISYS 752  
 QY 1377 VNY--DDTSKTSVYVVDVAVDDTIEVKKKLGKVTTLTSTGTGANKRALSQATGDL 1434  
 Db 753 GTVGGGQGHKLNLILD---NGTIV---KFLGDTTFNGCTKIEGKSIIQISNNYTTDH- 804  
 QY 1435 VKASDIVAHLNLTSGDIQAKAGSANSAGY-----VDADGKRVLYDSTDNKYYGA 1486  
 Db 805 VESADMTGTLFVNTPIYV---TLNKGAFVGVAKQYIISGPGN-IYFNEI----- 852  
 QY 1487 KNDGYDKTRKVAADKLVAAQOT-----PDGT-LAQNNAKSVINKEQVNDANKKOGIN 1538  
 Db 853 ---GNVGIYHGIANSISEFNASIGTSLFPGTPLDVLTKSTVNGTVDN-----FN 903  
 QY 1539 EDNAFVKGLEKASD-----NKTNAAYVVDLNAVAQPLPFGDTGTAKKLGELTLTI 1593  
 Db 904 APIVAVSGIDSMINQOIIIGDKKNIILISLSDNSITVANAULTYSGIRTKKNOG-TVVL 962  
 QY 1594 KGGQ-----TDNKLTD-----NNI-----GVVAGTD 1615  
 Db 963 SSGMPNPNPGTIYGLGLENSPKLYQVTTFTDINNLSGIANNVTTINDYVTLTTGGIAGTD 1022

QY 1616 GFTVKLAKDLTLMNSVNAAGTKIDKGVSEVDS-----SGQAKNPEVLSANGLD 1665  
 Db 1023 -PBAKI-----FLGSYNG-----NANVRFVSTSDPSMSVATVQANKGYTTY----- 1064  
 QY 1666 LGGRVTSNVGKGTQDDAANVOQLNEVRNLGLGNAGNDNADGNOVNIADIKKDPNSGSS 1725  
 Db 1065 LGNAIVSNI--GSLDTPVASVR-----FTGNDGSGAGLGONITSONID---FGT 1107  
 QY 1726 SNRTVAKAGTVLGR-----GNN-----DTEKATGVV-OVG 1756  
 Db 1108 YMLTILNSVVLGGGTFAINGEIDLTLNMLIPANGSTWGDNTSISTTLNVSNGNGGVV 1167  
 QY 1757 VDKDG-----NANGDLSNVWVKTKQDKSKKALLATVYNA--GQTYNLNN 1799  
 Db 1168 IAEADQVNAVTTGTTITKIQDANANFSGTQAVTILIOG-----ARENGTIGAPFEAVTG 1222  
 QY 1800 PAEADIRINEGIRFEHVNDQNEPVOGRNGIDSSAGKHSVAIGFOAKADGEAANAIG 1859  
 Db 1223 -----SNIFVKYELLRDSMODVYLFRINDVLNVVT-----1252  
 QY 1860 ROTQAGNOSIALGDNAQATGDO--STAIGTVNVACKHSGALGDPS-----TYKADNSY 1911  
 Db 1253 --TAVGNSAIA--NAPGVSONISRCLESTNTAAVNNMLAKDPDVAATFVGALATDTS 1307  
 QY 1912 SVG-----NNQPTDAMQDVFEGVGNNTVYESNSVALGSNSAISAGTHA-----GTQAK-- 1961  
 Db 1368 AVTYVLANDTQOTODLSNRGLTLEKLSAETSDVAGSATGAVSGDDAEVSYGVAKPF 1367  
 QY 1962 -----KSDGTAGTTTATAGATGVKG--FAGQTAGVAVSGAGABRRIOVNAAGEVS 2011  
 Db 1366 YNIAEDQKKKGIAQ--YRAKTTGVVVGDLTLASDMLIGAMIGITKIDIKHQDYKKGD-- 1423  
 QY 2012 ATSTDAVNG--SOLKAKAQIGAN-----ATNELDHRHONENKNAOISSAMMAS 2060  
 Db 1424 --KTD-INGLSLSLGSQOLYKNEFPAQGNALFTLKKVKSQORFFESNGMSKOIAGN 1480  
 QY 2061 MPQAYIPGRSVYTGIAITHNGGAVAY-----GLSKLSNGQ 2097  
 Db 1481 YDNMTF-GGNLIFG--YDYNAMPNVLVTPMAGLSTLKSNE 1518  
 RESULT 7  
 FHAB BORPE STANDARD; PRT; 3591 AA.  
 ID AC PI2255;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Filamentous hemagglutinin.  
 GN FHAB.  
 OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=90355839; PubMed=2388559.  
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RT "Genetic characterization of Bordetella pertussis filamentous  
 haemagglutinin: a protein processed from an unusually large  
 precursor."  
 RL Mol. Microbiol. 4:787-800(1990).  
 RN (2)  
 RP SEQUENCE OF 1-3261 FROM N.A.  
 RX MEDLINE=89202384; PubMed=2539596.  
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide  
 sequence and crucial role in adherence."  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).  
 CC -I- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND  
 INFECTION.  
 CC -I- SUBCELLULAR LOCATION: SURFACE.



Db 1785 GRIRA-----GDMHLDAPRIENTAKLSGEVQRKGVQDVGGEGHGRMSGIGYVWVWMLRA 1838  
 QY 1702 GNDMADNOVNINDIKDNDNSGS--SSNRTVIKAGTVL-----GKGNNTEKLTAGGVQV 1755  
 Db 1839 GNRKRGK---TIA---APWYGGDLTAEQSLIEVGKDLTLNAGAKDEHRLHNEGVIA 1891  
 QY 1756 GVDKDNANGDLSN----- 1769  
 Db 1892 G--GHGHIGGDVDNRSVFTVSAMEYFKTLPVSLALDNRAGLSPATWNPSTYELDY 1949  
 QY 1770 -----VW-----YKTKD----- 1777  
 Db 1950 LLDQNRXYEIMGLPYTTEMSVNTLKNLDLQYQAKPAPAPMPKAPBELDRLGHTLESAN 2009  
 QY 1778 -----GSKKALLATYNAAGOTVNLNPPAIDRIHQ--GIRPFYHNDGQBPVYQNRG 1831  
 Db 2010 GKRIEYKKLQGEYKAKMAVQAVEATRRVHDLQYKRGALGMDAETEVDGI 2069  
 QY 1832 IDSSASGKHSVAI--GFOAKADGEA-----AVAIGRTOAGNOSIAIG 1872  
 Db 2070 IGEFADLRTVYAKQADQATIDAEQDKVQRYKQSDAVRLQAIQPGRYTLAKLSALIG 2129  
 QY 1873 DNAQATGDSIALGICGNVYVAGKSHGAGIDPSTYKADNSTSVGNNOFTDATTDTVFCGN 1932  
 Db 2130 ADMRALGHSOLQMRWKDFKAGRGAEIA-----FYPKETQ--VLAAGA 2170  
 QY 1933 NITVESNSVALGNSAISAGTHAGTO--AKKSDGAGTATTAGATGYKVGAGOTAVGA 1990  
 Db 2171 GLTIS--NGAIHNGENMAQNRGPEGLIGAHSAISVSSGFDALRDVGLKRLDIDDALAA 2229  
 QY 1991 VSYG-----ASGAERRIONVAAGEVSAT-----STDAVNGSOLYKA 2026  
 Db 2230 VLVNPHITFRIGAAQTSGLADGAAPALARQARQAPETDGMVDARGLSADALASLASIDA 2289  
 QY 2027 TQGINATNELHRIHONENKANAGIS--SAMAMASMPQATIPGSMYTGCIATINOGA 2084  
 Db 2290 AAGL-----EVSGR--RNAQVADAGLAGPSAVALPAGVADV--GVEPTIG--DQVDPV 2338  
 QY 2085 VAVGLSK 2091  
 Db 2339 VAVGLEQ 2345  
 RESULT 8  
 HLYA\_PROMI STANDARD: PRF: 1577 AA.  
 ID HLYA\_PROMI STANDARD: PRF: 1577 AA.  
 AC P16466:  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Hemolysin precursor.  
 GN HPMA.  
 OS Proteus mirabilis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Proteus.  
 OX NCBI\_TaxID=584;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
 RC STRAIN-ISOLATE 477-12;  
 RX MEDLINE=90170827; PubMed=2407716;  
 RA Upjohn T.S., Welch R.A.;  
 RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin genes (hpmA and hpmB) reveals sequence similarity with the RT Seratia marcescens hemolysin genes (shlA and shlB).";  
 RL J. Bacteriol. 172:1206-1216(1990).  
 CC - FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
 CC - FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA REQUIRES HPMB FUNCTION.  
 CC - SUBCELLULAR LOCATION: Outer membrane.  
 CC - MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA

CC MAY BE RESPONSIBLE FOR PORE FORMATION.  
 CC - SIMILARITY: TO S. MARCESCENS HEMOLYSIN (SHLA).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M30186; AAA25657.1; .  
 DR PIR; A35140; A35140.  
 KW Hemolysis; Toxin; Outer membrane; Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1577 HEMOLYSIN  
 SO SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match 4.48; Score 474.5; DB 1; Length 1577;  
 Best Local Similarity 21.88; Pred. No. 8e-11;  
 Matches 397; Conservative 209; Mismatches 597; Indels 615; Gaps 96;

QY 506 KVDKNGDANDSKITTVGKNNKPDQOVNTLKLKGGVGVDTTETNGTVFGL----- 558  
 Db 7 KLSPSGRILASLALIFVSLNAYGNGIVDA-----GHGQPVSAVANGSTQVINITVTPNNEG 62  
 QY 559 ---NQ-----NGLTVGNSTLNNDGLSVKNTSNKQIQVYAGDGTFTDISNS 602  
 Db 63 ISHNQYQDFENVKPGFAVFNALFAGOSQL-----AGHLNANSNLNGQASLILNEVSR 116  
 QY 603 KP-----GAGIE-----NTRIRPDGIGFANNNGSLDANKPRLTPTGIANAGKELT 648  
 Db 117 NPSFLDQGEVFGIAAEVYLSNPMGITCQDGGFTNRS-----SLVYG----- 160  
 QY 649 NVQSAINPATNGQLDFPMNRLSTANTEKSGSAATIKDLYNLSOVPITFAGDTGPNYTKL 708  
 Db 161 -----NPLEFENGQLKGYSTLNTN-----LLSLCKNGLNTTGLL 194  
 QY 709 GEILK--VKGKTTADDLTKNNIGVAVDSTDNLSITVYLAFTLSDLDVAVT--KLTLSDKV 765  
 Db 195 DLAPRIDSRSKTIYAEIS-----AFTGQNTFS-----OHFDLSSQKPSALDSY 240  
 QY 766 TVDSGNNTAKLQNGDLFFSKQNTGATPATNSKITGVGDLKFTDNNGLADGTYITRKDY 825  
 Db 241 EFGS-----MQSRITITPAES-----SGVKLAG--KFTINDLSVADINQIDSOY 286  
 QY 826 GFARQDGLSKRPYLDKRLKVEVEITTINGINAGKATIGLNTLTATNATGHWYQ 885  
 Db 287 KYDSYD--KDGSENY--QVVRG--ITVNN-----SGSSQTLT--KTELKKNITL 329  
 QY 866 LGYDSTDKTRASI--GDVLNAGFNLNKNGDAKDPVSTDYDFINGNATTKAVT----- 939  
 Db 330 --VASSHNQIKASDLMDGDDTLTGADLT-----IDGQLOQKEKDIDN 370  
 QY 940 ---YDGRASRVAVYVNDVDTTILHTGADGKNOIGVY--TTTLTKTDKAGDAINFVS 993  
 Db 371 RWEFSWK---YDVYKEKQIQIG-----SGLDKANNATLTATGDTVLDA--KIN 417  
 QY 994 SGDDKALINKADIADINLTAGELIRNTKGTADTALOTFOYKRYKENGDDNDADITTVK 1053  
 Db 418 AGNNLAINAKDI--HINGLV-----EKESRSKNGKRNHRSLESGS 458  
 QY 1054 DAKNOVNTLKLK---GRN-GLDIQ-----TND 1078  
 Db 459 WSNHQTETLKASBLTAGKGLDAGOSTTAQAGAKLHANEVLVNAKDNINLNVQKTND 518  
 QY 1079 GTVFFGINTQSGKAGNNTLNNGLSIKNT-----AGNQIOVAGDGVK----- 1123  
 Db 519 KTYLDNRYVMGIGGQGNKN--NNNQOVSHAPLTADGQLLAADNNVNTISQVYKNG 577  
 QY 1124 -PAKVNNGV-----GAGIDSTRTIDDEIGFAGTNGS---LDSKSPH 1162



DB 578 AFVKTGGVVDNALSETISKIDERTGAFNITSSKHNKNTKOTSTGSELISDAQLTV 637  
 QY 1163 LSKDGINAGKRTNIQS-----GEIAONSNDVATGKIDKLENNKISSITAKTAQN 1216  
 DB 638 VAGNDVNVIGSLIKSADKLGISLADINVKSAQVLT--KIDDEKTSL--ALIGHAK----- 689  
 QY 1217 SLHESVADQ-----GNPF--VSNPSSVDTSKTSVITPAGNGTITVKNKGVAVGIDQ 1272  
 DB 690 -----EVEDKQYSAGFHITHTNKNTSTETEQANS--TISGAN-VDLQANKDVTPEAGSDL 741  
 QY 1273 TKGLETPKLVGNM--GKGIYI-----DSONGONTITG----- 1304  
 DB 742 -----KTTAGNASITGNVAFVSTENKQKQIDNTDITISGFSYTGVDKYGKADQY 794  
 QY 1305 -LSNTLANTNDKGSVTRTEQKIKIDEDKTRASTVLYSAGFN-----LQNGEAVD 1357  
 DB 795 DKHQTQTEVTKNRGS-QTEVAGDLTITANK-----DLHGHASHHVEGRQESGENIO 846  
 QY 1358 FVSTYDVFADGNATTAATVITDDTSKTSK-----VYDVNVDDTITLEVKDKL---GVKT 1410  
 DB 847 HLAVNDG-----ETSKTDSLNVGIDVGNLDYSGVTRKPKRAIEDGVNT 890  
 QY 1411 TLTSTGTGANKFALSNOATGALVASDVAHLNLTLS-----GDIOTAK 1455  
 DB 891 T-----KRCNNTDLTKVY-----ARDAIANLANLSHLETPNNGVEGIRGGSSQSQ 938  
 QY 1456 GASQANS-----SAGYVDADGNKYLYDSTDNKYQAKNDGTVDKTRKVAKDLYAQAQTPDG 1512  
 DB 939 TDSQAVSTINAGKIDIDSNKMLHD--QCTHYQSTQEGI-----SLANTHTSBA 986  
 QY 1513 TL-----AQMNVKSVINKEQVNDANKKOGIMEDNAFVKGLEKASDNKTKNA 1559  
 DB 987 TLKHQHTTHERKGGQIGVSTRTGSD--ITVAIKGGQTTDALMETKAKGQSTNSGDI 1045  
 QY 1560 AVTVGD-----LNAVAQTPLEFAGDT-----GATKAKLIGETLT 1593  
 DB 1046 SINVENAHTEGAQFPAOKGKIVYINAGDLTLAQTDTSEGSNNVNGSANKLVCTPPS 1105  
 QY 1594 K-----GGQDPTNK-LFDNNIGVAVGDTFTVLAKDLT-----NLNSVNA 1633  
 DB 1106 KDYGSGFNAGTTHHSKEQTAKVGTITGSQIEILNAGHMLTLQGHLSSEODIALNATN- 1164  
 QY 1634 GGRKIDDKGVSPVDSGQAKANPVLSANGDLGKVIYNVGCTPDTPAANVQULNEVR 1693  
 DB 1165 --KVLDQAS--SEHTEGN-----NLGSGVQAGFGKMTD--DASSVNGLSGAQ 1208  
 QY 1694 NLG-----LGAGNDNADGN-----QVNIADIKKDPNSG-----SSNRT 1729  
 DB 1209 FAIGKODEKSVREGGTINNSGMLTNGNSVHLQGAQVNSKDTQLTQSOGDIEITSAQST 1268  
 QY 1730 VIK--GYVLG--GKGNNDT-----EKLAT-----GGQVGVND-----KDNANGDLS 1768  
 DB 1269 DYKNNMGTDIGFNGKKTNTPKVEETKEKATSIHNIGKLLVNEVDOQKTSHQNATLENG 1328  
 QY 1769 NAWVKTKQD-----GSKKALLATYNAAGOTNYLTNNPAKIDIRINQGR--PFVHNDGN 1821  
 DB 1329 TLTINSKKDLTLISGANVTDVYGVNGGSLNLSQRES--DRHVTYGVAVGNNHNDPK 1385  
 QY 1822 QEPVQGRNIGDSSASG-----KHSVAIGFOAKADGEAAVAVIGROTAGNOSIA----- 1870  
 DB 1386 SSQVKN-----TAKAGGSLEKTIKOTIDSGIKSSPD-----AISDKYNSLSSTIADKIG 1435  
 QY 1871 IGDNAQATGDSIA-IGTG--NVYAGKHSAGIDPSTVADNSVSVGNNOFTDATQTDY 1927  
 DB 1436 ISDETKAKIDQGFKGKNGIKINIVTG-----AEGHTANADIKVTHVDNDAVTKT----- 1485  
 QY 1928 FGVGNNTTVESNSVALGNSAISAGHNAQTKAKSGTGCTTTGATGATCVGVGFGQTA 1987  
 DB 1486 -----SLTSSNDLSLNVN-----GSTKLTA--EIVSQGQOVD 1516  
 QY 1988 VQAVSVGASGAERRIQWY 2005  
 DB 1517 LGSSSV-----KLENI 1527

RESULT 9  
 OMPB\_RICCN STANDARD; PRT; 1655 AA.  
 ID OMPB\_RICCN  
 AC O9KKA3; O9KK98; O9XC45;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 01-MAR-2002 (Rel. 41; Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)  
 DE (OMP.B) [Contains: 120 kDa surface-exposed protein (surface protein  
 antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].  
 GN OMPB OR NC1085.  
 OS Rickettsia conorii.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CX NCBI\_Taxid=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 RN [2]  
 RP SEQUENCE OF 33-1649 FROM N.A.  
 RC STRAIN=Indian tick typhus, and Malish 7;  
 RX MEDLINE=20393643; PubMed=10939649;  
 RA Roux V., Raoult D.;  
 RT "Phylogenetic analysis of members of the genus Rickettsia using the  
 RT gene coding the outer-membrane protein ompb (ompb).";  
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).  
 RN [3]  
 RP SEQUENCE OF 353-1655 FROM N.A.  
 RC STRAIN=Malish 7;  
 RA Stenos J., Walker D.;  
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia  
 RT australis, the most divergent rickettsia of the spotted fever group";  
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AE008659; AAL03623.1;  
 CC EMBL: AF123721; AAF34124.1;  
 CC EMBL: AF123726; AAF34129.1;  
 CC EMBL: AF149110; AAD39533.1;  
 CC InterPro: IPR003858; OMPA\_rOMPb.  
 CC Pfam: PF02708; rOMPa\_rOMPb; 1.  
 CC Antigen: S-layer; Cell wall;  
 CC CHAIN: 1  
 CC CHAIN: 1335 1655  
 CC VARIANT 61 61  
 CC VARIANT 75 75  
 CC VARIANT 78 78  
 CC VARIANT 251 251  
 FT Complete proteome.  
 FT 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT 32 KDA BETA PEPTIDE  
 FT P -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT G -> S (IN STRAIN INDIAN TICK TYPHUS).  
 FT K -> N (IN STRAIN INDIAN TICK TYPHUS).  
 FT V -> A (IN STRAIN INDIAN TICK TYPHUS).

FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
 FT CONFLICT 353 354 KD -> GH (IN REF. 3).  
 FT CONFLICT 776 776 F -> S (IN REF. 3).  
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).  
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).  
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).  
 SO SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CAC64;

Query Match 4.3%; Score 456.5; DB 1; Length 1655;  
 Best Local Similarity 22.0%; Pred. No. 4e-10;  
 Matches 408; Conservative 191; Mismatches 677; Indels 579; Gaps 96;

QY 473 LTACVVDAAAT-----YGLKRVNOTABESALOTFFVKKVKKGGDAND---SKI 519  
 DB 12 ISAGLVASTATIVASFAGSAMGAIAIOQNRITNAVATTVGVGFQDQAVPANAVAPLANAV 71  
 QY 520 ITVGN-----NKPQOVNLTCLKGENGVDTETNGVTYFGLNONGSLTVGNSTLND 574  
 DB 72 ITAGVNGGIIINLPAGS-FNGLEPLTANNIDYVREDTTLGFIITVYVNAHFNLM--- 127  
 QY 575 GLSVKNTNSKQIQVADGITEFDISNKPAGIENTTRITRDGIGFANNTGSLDANKPR 634  
 DB 128 -----NAGKITLITGOGIT-----NVOAATKKNANN-----VAQVNGGAIDNND- 168  
 QY 635 LTPYG-INAGKELTYOASAINPATNGQIDFENRLSTANTEKSGSAATIKDLYNLSQVP 693  
 DB 169 LQVGRIDCGAASALVFENLANPT-----QKAP 197  
 QY 694 LTFAGDGPVTKLGEILKYKGGKTTADDLTNNIGVADS--TNSLTVKLAKTL- 748  
 DB 198 LI-----LGNAYIVGANGTLANVTNGFIKVSSEKSPATYVAVINIGGQIGMEN 245  
 QY 749 SDDAVNTKTLASDKVYVDSGNNTAKLQNGDLTFKQNTGATPATNSKITGYDGLKFTD 808  
 DB 246 TDADNVNTMLQANGATITFENGTD--GTGRVLKSKNAATDFNVTSGL- 293  
 QY 809 NNGALDGTYYITKDKYKFGKQSGSLDKSPRYDKDKLAKVEVEITTINGINAGKAITGL 868  
 DB 294 -----GNLK-----GIIIEFTVAVNGOLKANAG- 316  
 QY 869 SMTLTDATNATGHTVYOLGIVDSTDKTRAASI-GDYLNAGFNLKNGGADPFSTVDYD 927  
 DB 317 ANNAVIGTNNAGBRAA--GFVSVVDNGKATIDGOYV-----AKDW- 356  
 QY 928 FINGNATTAKYTYDGRASKAVAYVNDGTTILTGADGNKNOIGVKTTLTKTDAGDKRA 987  
 DB 357 -IGSANAVGVNF-----RHIVDVGDTGTAFKTA- 390  
 QY 988 INFSVNG--DDKALINAKDIADNLNLAGEIRTKGTADTALOTPOVKYKENG---DDD 1043  
 DB 391 ITONSNEGTDFENLAQIIVPTMTLNG--NFTGDASNPENAGVITFEDAGTLASAS 447  
 QY 1044 NDADITTVGKADATNOYNTLKLKGNGLDIQTKDGTVEFGITOSGLKAGNNTTLNNG 1103  
 DB 448 ADANV-----AVTNNTITAEASGAGVYOLSTHAEMLG-NAGSVFLADGTYVNGR- 499  
 QY 1104 LSKTNAGNEOIOVADGVKFAKVNNGVAG--IDGTTTRITRDEIGFAGTNGSLDKS 1160  
 DB 500 -----VNOTALVG-----GALAGATTLDGSATITGD-IGNAGGAALOGIT 540  
 QY 1161 PHLKSG--INAGKKTINISGELIAGNSNDNAVYGGKIYDKLTELEKISSATAKONS 1217  
 DB 541 --LANDATKTLTGANITIGANGTL-----NFQANGGT-----KLTST- 581  
 QY 1218 L--HEFSVADEQGNFTVSNPYSDTSTKSDVTFPAGENGITTVKNGGVAVGIDOK 1274  
 DB 582 IYVDFLALATDQ--TGVDASSLTNAQT--LTINKIGTVGANKTKLQFNIGSSK 634  
 QY 1275 GLTTP-----KLTGVNNN-----GKG-IVIDSQNGQNTITGLSMTL 1309

DB 635 TVISDGDVAINEIVYNNAGVAFANHTYLLITRTNAGGCKIIFNPVYNNNTTLATGTL 694  
 QY 1310 ANVTNDKGSVRYTEBOCKIIRKEDKTRAASIVDVLASGFNLOGNGRAVDFVSTYDV- 1365  
 DB 695 GSATNPLAEINFGSK-----AANVDIVLANGVK- NLVATNTTTTDAVGSFI 742  
 QY 1306 ANADGNATTAKYTYDTSKTSKVYDVAVDOTTLEVKDKKLGVT-----TTLSTGT- 1418  
 DB 743 ENAGGTIVSGVGGQGGKRFNTVALD--NGTIV--KFLGNATFNGNTTIANSTLQ 795  
 QY 1419 -GAN--KFLSNQATGDALVAKASDIVAHLNLTSGDIOTAKASQASAGV--VADNCK 1473  
 DB 796 IGGNTIADVAASADTG-----IVEVNTGPITVTLNKAAVYVNAKQITVSGPVG 847  
 QY 1474 VIYDSTDNKYQAKN-DGYDRTKEVAKDKLVQAOTP-----DGLTQANVKSIVNK 1525  
 DB 848 VI-----NEIGNAGNVHGAVTDTIAFENSLGAVFELPGIFENDAGRIPLTIKSTVGN 902  
 QY 1526 EOVNDANKKOGINEDNAFYKLEKASDNKT--KNAAVTVG--DLNAVQTPLEFGD 1579  
 DB 903 KAT-----GFDVBSVIVLGVDSVYADGVYIGDONNTVGLGSDNDIIVATTLVAG- 955  
 QY 1580 FGTTAKLGETLTIKGGQTD-----NKLPD--NNIGVAGT----- 1614  
 DB 956 IGTINNNG--TYTLSSGIPNTPGYVGLCTGIGASKFKQVFTTIDYNNLGNITATIN 1014  
 QY 1615 DGEVYKLA-----KDLTNLSVNAAGTKIDDKGVSEVDS-----SQOANRPEVL 1659  
 DB 1015 DGVTVTTGIGIAGIDFGDKITLISVNGC-----NVRRVGDILSHSTSMIGTANNQTV 1068  
 QY 1660 SANGDLGKVISNNGKGTGFODANVOOLNEVRLTLGAGNNDNAGNOVNIADIKD 1719  
 DB 1069 TY-----LGNAFVGNF--GSDTPVASVR-----FTSGGAGLOGINTYSOYID 1110  
 QY 1720 ---PNSGSSNRFVIKAG--TVLGGKNDTEKLATGVQGVDRKGNAGDLSNVWYVQ 1775  
 DB 1111 FGTYNLGSNSVILLGGTITAINKINLRNTLT-----ASG--TSTW- 1152  
 QY 1776 KDSKKALAIYNAAGOTNYLTNPAAEAIIDRINEGIRFFHVNDGNDEPVVGRNGIDSS 1835  
 DB 1153 --GNNTSIETTLTLLA-----NGNIGNIV----- 1173  
 QY 1836 ASGKSHVAIGFOAKADGEAAVAGROQAGNOSTAIGNQA--TGQSLAIGTGVYAG 1893  
 DB 1174 -----ILEGQVNA-----TTGTTTAKVODNANAFSGTQTYTLIG--GA 1213  
 QY 1894 KHSAGIDPSTV-----KADNSYSVNNNOFTDATOYDF-----GVG 1931  
 DB 1214 RENGTLGGNPFVYVTSNRFVYVGLIRANODYVITRTNNAENVYVINDIANSFGGAGVG 1273  
 QY 1932 NNIT--VIESNSVA-----LSNSAISAG-----THAGTOAKKSDGTACTT 1971  
 DB 1274 QVTFVNAVTNAAVNNLLKANSANSANFVAGVITDTSAITNAOLDVAK-DIOADLGN 1332  
 QY 1972 TAGA-----TGTVKFACTAVGAVSVGASGAERIORVAG-BVASTSDA- 2021  
 DB 1333 RIGALRYLGTEPAMAPPEA-GALPAVVAAGDERVDAVWAGIWAKEPYTDAHOSKGL 1391  
 QY 2022 QLYKA--TOGINAATNELDRITHONENKANAGISSAMASMP--QAYIPGRSVWTG- 2075  
 DB 1392 AGYKAKTGVYIGDPL-----ANDNLM--IGAIGITTKIDKHQYKKGDKTDVNGFS 1443  
 QY 2076 -----IATHNGQAVANGSLKSLDNGOVWRKINGSADTQGHVAAVAGFH 2121  
 DB 1444 FSLGAQOVLVNFPAOGSAIFSLNQVKNKSQRYF-----FDANGMMSKQIAAGHY 1493

RESULT: 10  
 P3P\_LACLC  
 ID P3P\_LACLC STANDARD; PRT; 1902 AA.  
 AC P15292;  
 DT 01-Apr-1990 (Rel. 14, Created)

01-APR-1990 (Rel. 14, last sequence update)  
 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Pili-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
 associated serine proteinase).  
 GN Pili-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
 associated serine proteinase).  
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
 OC plasmid.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1359;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.  
 RC STRAIN-SK11;  
 RX MEDLINE=89340435; PubMed=2760036;  
 RA Vos P., Simons G., Siezen R.J., de Vos W.M.;  
 RT "Primary structure and organization of the gene for a procaryotic,  
 cell envelope-located serine proteinase.";  
 RL J. Biol. Chem. 264:13579-13585(1989).  
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 GROWTH OF THE BACTERIA ON MILK.  
 CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
 specificity, although some substrate preference have been noted,  
 e.g. large hydrophobic residues in the P1 and P4 positions, and  
 pro in the P2 position. Best known for its action on caseins,  
 although it has been shown to hydrolyse hemoglobin and oxidized  
 insulin B-chain.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 SUBTILASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL J04962; AAA03533.1; ALT\_SEQ.  
 DR PIR: A32634; A32634.  
 DR HSSP: P00782; 2SBR.  
 DR MEROPS: S08.019; -;  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR003337; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF02225; PA; 1.  
 DR Pfam: PF00082; Peptidase\_S8; 3.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.  
 DR Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;  
 KW Transmembrane.  
 KM  
 FT SIGNAL 1 33  
 FT PROPEP 34 187  
 FT CHAIN 188 1902  
 FT DOMAIN 188 1876  
 FT TRANSMEM 1877 1895  
 FT DOMAIN 1896 1902  
 FT ACT\_SITE 217 217  
 FT ACT\_SITE 281 281  
 FT ACT\_SITE 620 620  
 FT DOMAIN 1867 1872  
 FT  
 SO SEQUENCE 1902 AA; 200550 MW; 87CEBBA9345F9D3 CRC64;

Query Match 4.2%; Score 444.5; DB 1; Length 1902;  
 Best Local Similarity 20.8%; Pred. No. 1.3e-09;  
 Matches 443; Conservative 219; Mismatches 753; Indels 715; Gaps 100;

QY 370 RVEKGLKLTITGGAQTSALTNDHNVGNGDLKVQ-----LAETLT--SLKAVTTEN 422  
 Db 370 RVEKGLKLTITGGAQTSALTNDHNVGNGDLKVQ-----LAETLT--SLKAVTTEN 422  
 QY 423 RKKKGL--SILLAGTYALGLAVLPVGEIQAARAIISQIKGSSLANVYAAATKQAATPE 60  
 Db 423 RKKKGL--SILLAGTYALGLAVLPVGEIQAARAIISQIKGSSLANVYAAATKQAATPE 60  
 QY 423 LTNKRYVAGTRITLDKIGFTNDMGIDSKPYLD----- 458  
 Db 423 LTNKRYVAGTRITLDKIGFTNDMGIDSKPYLD----- 458  
 QY 459 -----KDTGHAGGOKITKLAVDDAATYG-----QLKKVNGT 494  
 Db 459 -----KDTGHAGGOKITKLAVDDAATYG-----QLKKVNGT 494  
 QY 119 EIQDETNRKVIQAASVYAAAEVYQQTAG-----ESYGVVNGFSTKYRVYDIPKQI 172  
 Db 119 EIQDETNRKVIQAASVYAAAEVYQQTAG-----ESYGVVNGFSTKYRVYDIPKQI 172  
 QY 495 AESALQETVYKVV-----DKNGANDSKITTYGKNNKPDGTQVNTLKGENGVDVT--- 547  
 Db 495 AESALQETVYKVV-----DKNGANDSKITTYGKNNKPDGTQVNTLKGENGVDVT--- 547  
 QY 173 A--GVKTVTLAKYYPIDPAANSANVQAVMSKTKGETVYSV---DSGIDPTKHD 226  
 Db 173 A--GVKTVTLAKYYPIDPAANSANVQAVMSKTKGETVYSV---DSGIDPTKHD 226  
 QY 548 -----TET-----NGYTFGLN--QNN----- 562  
 Db 548 -----TET-----NGYTFGLN--QNN----- 562  
 QY 227 MRLSDKDVKLTKSDVEKFTDVGKRYFNSKYPGFYGNNDITITDDKVDEQHGMYA 286  
 Db 227 MRLSDKDVKLTKSDVEKFTDVGKRYFNSKYPGFYGNNDITITDDKVDEQHGMYA 286  
 QY 563 GLTVGNSLTNNDGSLVNTNSKQI---QVADGIFTTDSNSKPGAGIENTRITRD-- 617  
 Db 563 GLTVGNSLTNNDGSLVNTNSKQI---QVADGIFTTDSNSKPGAGIENTRITRD-- 617  
 QY 287 GIIANGTGDDPAKSVGVAPAEQILAMKYFNSDTSKTSATVVSALDSKIGADVL 346  
 Db 287 GIIANGTGDDPAKSVGVAPAEQILAMKYFNSDTSKTSATVVSALDSKIGADVL 346  
 QY 618 GIGFANNGLSILANKPRLTPTGIMAGKELTNVOSAINPATNGOLDPMNR----- 668  
 Db 618 GIGFANNGLSILANKPRLTPTGIMAGKELTNVOSAINPATNGOLDPMNR----- 668  
 QY 347 NMSLGSSNGOTLEDPELAIVQ--NANSGTAAYIASGNSGTSATGKVDYGLQDNE 405  
 Db 347 NMSLGSSNGOTLEDPELAIVQ--NANSGTAAYIASGNSGTSATGKVDYGLQDNE 405  
 QY 669 -----LSTANERKSGSAITKDYNLGCVPLT-----FAG----- 698  
 Db 669 -----LSTANERKSGSAITKDYNLGCVPLT-----FAG----- 698  
 QY 406 MVSFGTSRGATVVAASENDVITQAVTITDGTGLQGPETIQLSHDFGSPDKKFEYI 465  
 Db 406 MVSFGTSRGATVVAASENDVITQAVTITDGTGLQGPETIQLSHDFGSPDKKFEYI 465  
 QY 699 --DGPVNTK-----KIGELTKYKGGKTITDGLK-----NNGVAVSDNSLT- 741  
 Db 699 --DGPVNTK-----KIGELTKYKGGKTITDGLK-----NNGVAVSDNSLT- 741  
 QY 466 VKDSGNSKSLALADYADAKGKTAIVKRGFSFDDKQKTAQAAGAGLIIVNDGTATP 525  
 Db 466 VKDSGNSKSLALADYADAKGKTAIVKRGFSFDDKQKTAQAAGAGLIIVNDGTATP 525  
 QY 742 --VKLAKTIS--DLDVNTKLTASDKVYVDSGNN-----TAKLQNGDLTESKON--T 788  
 Db 742 --VKLAKTIS--DLDVNTKLTASDKVYVDSGNN-----TAKLQNGDLTESKON--T 788  
 QY 526 MTSIALTTTPTTPELSSVTGQKLV--DWTVAHPDPSLGVKTLTAMLNQYTEDKMSDFT 583  
 Db 526 MTSIALTTTPTTPELSSVTGQKLV--DWTVAHPDPSLGVKTLTAMLNQYTEDKMSDFT 583  
 QY 789 GATPATNSKTIQVGLKFTDNNGIALDGTITYTKDYKFAKQDGL----- 834  
 Db 789 GATPATNSKTIQVGLKFTDNNGIALDGTITYTKDYKFAKQDGL----- 834  
 QY 584 SYGPVSN-----LSFKPDI--TAPGGINIWSSTONNNGYTNMSGTSNASPIAGSQALLK 634  
 Db 584 SYGPVSN-----LSFKPDI--TAPGGINIWSSTONNNGYTNMSGTSNASPIAGSQALLK 634  
 QY 835 -----DKSKPYLDKDKLNG-----EVEITL-----NGINAGKALTGSLTITDPTA 878  
 Db 835 -----DKSKPYLDKDKLNG-----EVEITL-----NGINAGKALTGSLTITDPTA 878  
 QY 635 QALNKNKPFYAYYKQKLTALTPLEKTVEMNTAQPIINDIN-----YANVVSRRQ 686  
 Db 635 QALNKNKPFYAYYKQKLTALTPLEKTVEMNTAQPIINDIN-----YANVVSRRQ 686  
 QY 879 TTGHVTOGLIVSDTKTRFASIGDVNLNAGFNKNGN-----DAKDFVSTYDVFINGNAT 934  
 Db 879 TTGHVTOGLIVSDTKTRFASIGDVNLNAGFNKNGN-----DAKDFVSTYDVFINGNAT 934  
 QY 687 GAGLVDVAAALDALEKNSTYVAE-----NGYPVAVELKQFTSDKTEKLTFTNRT 736  
 Db 687 GAGLVDVAAALDALEKNSTYVAE-----NGYPVAVELKQFTSDKTEKLTFTNRT 736  
 QY 935 TAKVTV-----DGKA-----SKYAVDVNVNDGTTI----- 958  
 Db 935 TAKVTV-----DGKA-----SKYAVDVNVNDGTTI----- 958  
 QY 737 THTLYQMDSNITNVAVYSATDPNSGVLYDKIDGAIKAGSNITVPAGKTAQIEFTLS 796  
 Db 737 THTLYQMDSNITNVAVYSATDPNSGVLYDKIDGAIKAGSNITVPAGKTAQIEFTLS 796  
 QY 959 -----HLTGADGNKNOI-----GVK-----TT 975  
 Db 959 -----HLTGADGNKNOI-----GVK-----TT 975  
 QY 797 LPSKTDQOOEYBGLFNRGSDGSRNLNLYMGFFGDMDNGKLYVDSLNGITTPYPAGNGFTV 856  
 Db 797 LPSKTDQOOEYBGLFNRGSDGSRNLNLYMGFFGDMDNGKLYVDSLNGITTPYPAGNGFTV 856  
 QY 976 TLTK-----TDAG-----DKAINFSVSGDDKALIN----- 1002  
 Db 976 TLTK-----TDAG-----DKAINFSVSGDDKALIN----- 1002  
 QY 857 PLTKKKNKTGYQYGGVAVDADGNKTVDDQALAF-----SSDKNALVNDISMKYLLRNISNV 913  
 Db 857 PLTKKKNKTGYQYGGVAVDADGNKTVDDQALAF-----SSDKNALVNDISMKYLLRNISNV 913  
 QY 1003 AKDIAD-----NNTLAGIRNTK-----GTADPTAL 1028  
 Db 1003 AKDIAD-----NNTLAGIRNTK-----GTADPTAL 1028  
 QY 914 QVDIILDGQGNKVTTLSSSTNRKTYYNHNSOQYIYYNAPAWDGTYYDQDGNIKITADGS 973  
 Db 914 QVDIILDGQGNKVTTLSSSTNRKTYYNHNSOQYIYYNAPAWDGTYYDQDGNIKITADGS 973  
 QY 1029 GTFQVKKYKENGDD-----DNDADTIT--VGKDAKTQVNT--LKLKGR---NGL 1071  
 Db 1029 GTFQVKKYKENGDD-----DNDADTIT--VGKDAKTQVNT--LKLKGR---NGL 1071  
 QY 974 YTYRISGVPEGSDKQVQVDFPKDKSKAPYRHAVALSKTENGTQYLLTEAADDLSGL 1033  
 Db 974 YTYRISGVPEGSDKQVQVDFPKDKSKAPYRHAVALSKTENGTQYLLTEAADDLSGL 1033  
 QY 1072 DIO-----TNKDKIVTFGINTQSLGAKAGNNTTLNNGSLIKTAGNEQIOVGADG 1121  
 Db 1072 DIO-----TNKDKIVTFGINTQSLGAKAGNNTTLNNGSLIKTAGNEQIOVGADG 1121

```

Db 1034 DATKSVKTEINEVTNDAFTDAGTADGY-----TKIETPLSDBOAOA----- 1077
QY 1122 VKFAKNNNGVVGACIDGCTRIRIRDELGEFAGTNGSLDKSPHLKSGC--INAG--KKTIN 1177
Db 1078 -----LGNG-----DENSELYITDASNAATODDASVQPGSTSPDLVINGGIDPKTSS 1126
QY 1178 10SGEIAONSNDNAVGTGKTYDKTELENKISSAKTQNSLHEFSVA--DEQNNFTVSNP 1236
Db 1127 TTGTGYEANTQ-----GGGYTFSGTYPAVVDGTYDQCKKDDMLNTTYAALNSFTASMP 1181
QY 1237 YSSYDTSKISDVITTFAGENGIT---TTKYNKGV-----VAVGIDQTKGLT----- 1277
Db 1182 VYMADYAAOVLDYADKAHTQLKHPDTKYRLMAPTTDLKFNNNGSDQSEATIKYGVYS 1241
QY 1278 --TPKLTGNNNGKGIIVDSONGONTTGLS--NTLANVTNDKDSVTRTECGKIINKED 1332
Db 1242 ADTKYVNVGHYTA--ALDAQHHSYDVYVNGDNTIKYATDKDQONTTECKTTSYD 1298
QY 1333 KTRAASTVDVLASGFNLQNGEAVDFVSTYDT--VNEFADG--NATTAKYVYDQTSKSYV 1389
Db 1299 P-----DMLKSV-----TFDQGVKFGTKKFNATSAKF--YDPKTIATIT 1337
QY 1390 YQVNVDDTITIEYKDKLGVK-----TTTILTSTGTGANKFAL-----SNOATGALVKAS 1438
Db 1338 GKVKHPTTLQYDQKQIPKIDDLTFSTLDLGLQKPGVYVGDPTQTKTFOEALSTFL 1397
QY 1439 DIYVHLNLTSGDIOTPAKGSQANSAGYVDADGNKVTYDSTDNKYKQAKNDQVTKREV 1498
Db 1398 DAVA--PTLSLDSST-----DAPVYINDPNEFOITGAIIDMAOTLS----- 1435
QY 1499 AKDKLVAQOTPDGTLAONNVKSVINK--EQVNDANKKQG--INEDNAFVKG--LEKAA 1551
Db 1436 -----LSINGSSVASQYEDININSGKPGHMAIDQPVKLECKANLVYAV 1479
QY 1552 SONKTNAAVYVGDNLNAVQOTPLTFAGDTGTAKKLGELITTKGQOTDNKTNNIGVY 1611
Db 1480 TQSEDN--TTTKNITVYEEPKTLAAPVTPS-----TTEPQOYV--LTAN----- 1522
QY 1612 ACTDGTGVLAKDLNLSNVNAGTKRID--KGVSVYDSSGQAKANPVLNSAGLGLG 1668
Db 1523 AATGTEVVOYSD-----GKTYQDVPAAGVT--ITANGTEFKSTDLTGNSPAVD 1572
QY 1669 KYISNVGKGTJDTDAANVOOL--NEVRNLLGLGNAGNDNAGNVNIADIKDPNSGSSN 1727
Db 1573 YVVTNI--KADDEPAQLOAKOELTNL-----IASAKTILASGKXYDD 1611
QY 1728 RTVIRAGTVLGKGNNDDEKATG--GVQVGYDKRGNANGDLSNVWVTKQDGSKKALLA 1785
Db 1612 ATT-----TALAANA--TOKAQOTALDQTNASVDSLGTANRDLOT-----ATNQLAA 1654
QY 1786 TYNAGQOTNYLTNPFAEIDRINEQIRFEFHNDGNEPVOGRNGIDSSASGKHSVAIG 1845
Db 1655 KLPAPAKKTSIL-----NQLOSKYDALGTLGQNOTDPSSTCKT 1690
QY 1846 FOAKKADGEAANAIGROTQAGNOSIAIGNAQTGDSIAIGTGVNVAKHSGAIGDPSSTV 1905
Db 1691 FTAALDDIYVA-----QAQAGTQT--DDLOQAT--LAKILDEVLAKIAGS-----I 1731
QY 1906 KADNSYSGVNNQFTDQOTDVEFGVNNITVYESVVALGSN-----SALSAG 1933
Db 1732 KATPAEYGNMAK--DAATGKTWADIADITLISQASADASDKLAHLQALOSLKTQVAAA 1788
QY 1954 THAGTQAKKSDTGTAGTTAGATGVKGFACOTAVGAVSVG--ASGAERRIONVAAGEV 2010
Db 1789 VEAATVVGKGDGTITSDKGGGQGPAPAPBDTGKDKDEDSQSPSSGG--NIPKPAITTT 1846
QY 2011 SATSTDAVNGSOLYKATGCIANATNELDHR 2040
Db 1847 STTDDTTDRNGOLTSKGALPKTGETTER 1876

```

RESULT 11

```

PIP_LACLC
ID PIP_LACLC STANDARD; PRT; 1902 AA.
AC P15271.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE proteinase).
GN PRT.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Plasmid pMV05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN 11.
RP SEQUENCE FROM N.A.
RC STRAIN=MG2.
RX MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris Wg2."
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC P2 IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24767; AA01677.1; -.
DR HSBP; P00782; 1S01.
DR MEROPS; S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR002029; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00882; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SO SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5DOB CRC64;

```

Query Match

4.1%; Score 437.5; DB 1; Length 1902;

Best Local Similarity 21.1%; Pred. No. 2.4e-09;  
Matches 453; Conservative 197; Mismatches 743; Indels 759; Gaps 100;

Oy 370 RVEKGLGTLITIGGAGTSAITDHNIGVYQNGDGLKVO-----LAETLT--SLKMTTEN 422  
 Db 3 RKKKGL--SILLAGVALGALAVLPVGEIOAKAAISOOTKSSLANVTAATAKQAADT 60  
 Oy 423 LTANKEVVGKTRLTDDIGFTNMNGIDESKPYLD----- 458  
 Db 61 TAAITNMAIA-TOLAKKIDY-KLKNKVOQODIYDVIVQSAAPASENGTLRFDYSSA 118  
 Oy 459-----KDTGHAAGOKITKLPAGVDDDAATYG-----OLKKNVOT 494  
 Db 119 ELQOETNKNVIAQAQSVKAAVEQVTOGTAG-----ESGYVNGPSTKVRVVDIPKLKQI 172  
 Oy 495 AESALQTTTAVK---DKNGDANDSKITIVGKNNKPDGQVNTLKIKENGVDVT--- 547  
 Db 173 A--GVKVTTLAKVYVPTDAKANSMAVQAVMSNYKYKGEQTVSVI---DSGIDPETHD 226  
 Oy 548-----TEP-----NGTVPEGLN--QNN----- 562  
 Db 227 MRLSDKQVYKLTLSDEVEFTDPAKHGRFNSKVPYGFNYADNNDITDDIYDEQGHMVA 286  
 Oy 563 GLTVGNSTLNDGLSVKNTNSNKOI---QVGADGITTIDISNRKGAIENTTTRTD-- 617  
 Db 287 GIGANGTGDPAKSVGVAPAEQOLAKKVPNTSDTSATGSSSTLYSAIEDSAKIGADVL 346  
 Oy 618 GIGFNNNTGSLDANKPRLTPGINAGKELTNVOSAINPATNGOLDPNR----- 668  
 Db 347 NMSLSDSGNOLTEDEPDLAAVO--NAMESGTAAVISAGNSGTSGSATEGVNNDYVGLADNE 405  
 Oy 669-----LSTANTEKSSAATIKDLYNLSQVPLT-----FAG----- 698  
 Db 406 MWTGTGSRGATTVASAENTDVIITQAVTITDGTGLQGPHTIOLSSNDFTSFPOKRTYV 465  
 Oy 699--DTGPNYTK-----KLGEILKVGKGTADDLTK-----NNIGVAD 734  
 Db 466 VYDASGNLSKGLADYTAADAKGIALYARGELSPDDKQKVAQAAGALIIYNNDGITP 525  
 Oy 735 ST-----DNLSTVLKATL-----SDLDV 754  
 Db 526 VTSMAITTTTFFGLSSVTGOKLYDMVTAHPDDSLGVKIALTLVYNQKTEKMDFTSY 585  
 Oy 755 N-TKLITSDKTYVDSGNNTAKLQNGDLTFKQNTGATP-ATNKTITGVDLAKFTDNGI 812  
 Db 566 GPVSNLSKRPDITAPGAINWSTQNNNGYTNNSGTSMASPIAGSALLKQALNNKNPFY 645  
 Oy 813 A-----LDGTYITTKDVGFAKODGSLDKSPYLDKDKLVEVEITV---NGINAGKA 864  
 Db 646 AYTKQLKGTAL-----TDFLKT--VENMTAQPIINDIN----- 675  
 Oy 865 ITGLSNTLTDAITNATGHTVQGLIVSDTKTFRASIGVLANGFNLKNG---DAKDFV 920  
 Db 676---YNNVIVSPRQAGLVADYKAALDALEKNPSTVAB-----NGPVELKDFE 722  
 Oy 921 STYDVIDFNGNATTAKYT---DGKA-----SKAYVYVNDGTTI 958  
 Db 723 STDKTFKLTFTSTHETLYQMSNTDITNAVYTSATDPNSGVLDKDKIDGAIRAGSNIT 782  
 Oy 959-----HLTGADG----- 965  
 Db 783 VPAGTAQIEFTLSLPKSFDDQQFVEGFLNFKSGDSRLNLPYMGFPDQMDNGKIVDSL 842  
 Oy 966-----NKNOIGVKTTLTKTTDAK---DKAINFVSNGDDKALIN 1002  
 Db 843 GITYSPAGNFGTVPLTNKN-TGTQYGGVMTDADGQGTVDQAIATF---SSDKNALYN 898  
 Oy 1003-----AKDIAD--NLMNTAGEIRNTR- 1021  
 Db 899 DISMRYILLRNISNVQVLLDQGNKVTLLSSSTNLTKTYNAHSQOXYIYNAPAMDGY 958  
 Oy 1022-----GTADTALQTEOVKKVKNENGDD-----DNDADITV--VGDAKATNQVNT 1062

Db 959 YDQRDGNIKRAGDSSTYTRISRGVPEGGDKRQVDPFKLDSKAPTVRHVALSAKTENGRT 1018  
 Oy 1063---LKLKGR---NGLDIQ-----TNKDGTVFEGINTGSLKAGNNTLNNGLSI 1106  
 Db 1019 QYTLTAEKDLSGLDITKSVKTAINEVTLMDITFTDAGTTADGY-----TKI 1066  
 Oy 1107 KNTAGNEQIUGADGVYFAKVNNGVAGIDGTTTRITREIGFAGTNGSLDKSRPHLSD 1166  
 Db 1067 EPLSDQAOA-----LGNG-----DNSAELIYDMSANATDQASVQKPGSTSE 1111  
 Oy 1167 G--IMAG--KKTINIOSGEIAONSNDATGKIYDLKTELEKISTAKTQNSIHEFS 1222  
 Db 1112 PLIYNGGIGIPDKISSITTEGYEANTQ-----GGGTYFSGTYPAAYGVITYNAQKKHHDIN 1166  
 Oy 1223 VA-DEGNNTFTVNPYSSYDTSTKSDIITFAGENG---TTKVNKGV-----VRGI 1270  
 Db 1167 TTYDAATNSFTASMPVYNADYAAOVLDYADKHAHTOLKHEFTVRLPAPFTDLKNNQS 1226  
 Oy 1271 DQFKGLT-----TPRLTVGNNGKGLVIDSON-----GONTITGLSNTLA 1310  
 Db 1227 DQISEATIKYTVGSADTKVNVGDTVA---ALDAQHFEVDVRYVNGDNTIK-----V 1277  
 Oy 1311 NVTNDKGSVTEGQKIKDEKTRAASIVDVLSAGFNLOGCEAVDPSTYTVFPAQ 1370  
 Db 1278 TATEDDENT--TTBQKITSSYDP-----DMLK-----NSVTFDQGVTFGANEF--- 1319  
 Oy 1371 NATTAKYTDISTKSVVYDVAVDOTTIEVKDKLGK---TTTLITSTGANKAPAL 1425  
 Db 1320 NATSAKF-YDPKGTIAITRKVKNHPTTLQVDSKQDIPKIDLTFFSLDGLTGQKPFV 1378  
 Oy 1426-----SNOATGVALVYASDPIVAHLNLTSGDIQTAQASQANSAGVYDAGKVIYDST 1479  
 Db 1379 VVGDTTQNKTFQELFTLIDAVA--PLLSLEST-----DAPYTNPNQIIGTAT 1428  
 Oy 1480 DN-KYQAKNDGT---VDKTEVAKDKLVAQADPTDGTLOAMVKSIVNKEQVNDAN 1532  
 Db 1429 DMAQYLSLSINGSSVASQYVDININSKPGHMAIDQ-PVKLLEGKRVLV---AVTD-- 1481  
 Oy 1533 KKGINDNAFVGLERAADNKTNAAYVYVDLNAVQPIPLFAGDTGTAKKLEETLT 1592  
 Db 1482-----SEDNNTTKNITVYEPKRTL-AAPIV-----STTEPAKTVTLT 1520  
 Oy 1593 IKGGQTDNKLTDNNIGVAGTDEFTYKLAKDILNLSVNAAGTKIDD---KGVSEVDS 1649  
 Db 1521 ANSAT-----GETVQYASD-----GKTYDVAAGVT--VTAN 1553  
 Oy 1650 GOAKANTPVLANGLDGKVIYSNVSGKGTDPDAANYQOL-NEVRNLLGIGNAGNDADG 1708  
 Db 1554 GTFKFKSTDLYGNSPAVDVYVNI---KADPPAQLOAKKOELTNL----- 1596  
 Oy 1709 NOVNIADIKKDPNSGSSNRTVYAGIVLGKGNDEKATG--GYQVGVDDKGNANGD 1766  
 Db 1597---IASAKTLSASGKYDDMT---TALAAA---TQKQOTLDDOTNASVDSLGTGNRD 1644  
 Oy 1767 LSNVWVYQKQKSKALLATYNAAGOTNYVLNMPAEALIDRINBQIGFFHVNDGNQEPV 1826  
 Db 1645 LOT-----ATNOJLAAPKIPAKTISL-----NOLQSV 1671  
 Oy 1827 QGRNGIDSASGKSHVAIGQAKADGEAAVAIGRQTAQNGQSIAGMAQATGDSQTAIG 1886  
 Db 1672 KAALGTDLQNGTDBSTGKTFTALDLVA-----QAQAGTQOT--DDQHOAT--LAKV 1719  
 Oy 1887 TGNVYAGKHSALIDPSTYKADNSYSVGNNOPTDATOTVDFVGNNTIYTESNSVALGS 1946  
 Db 1720 LDAVIAKLAEG---IKAAQPAEYGNAK---DAATGTWADIAIDLTLTSGQASDAS 1769  
 Oy 1947 N-----SAISAGTHAGTQAQKSDGTAGTTTAAAGAT---VKGFAGOTAVGAV 1991  
 Db 1770 DKLALHLOLSLKTVAANAFAATVYKGDGTTTSPKGGGQGTAPAPAPDIDGKDGDEG 1829  
 Oy 1992 SVGASGAEIRIONVAGEVSATST---DAVNGSOLYKATOGIANATNELDHR 2040  
 Db 1830 SQPSSG---NIPNPAITTTSTJSTDDTDRNGQLNSGKALPKGTGETTER 1876

RESULT 12  
P2P\_LACLC STANDARD; PRT; 1902 AA.  
ID P2P\_LACLC  
AC P13293;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-  
associated serine proteinase) (LP151).  
GN PRT.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OC Plasmid pLP763.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Lactococcus  
OC NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCDO 763;  
RX MEDLINE=89313288; PubMed=2501630;  
RA Kikuchi M., Ikemura H., Shimizu-Kadota M., Hirashima A.:  
RT "Molecular characterization of a cell wall-associated proteinase gene  
from Streptococcus lactis NCDO763.";  
RL Mol. Microbiol. 3:359-369(1989).  
CC -1- FUNCTION: PROTEINASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
GROWTH OF THE BACTERIA ON MILK  
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
specificity, although some substrate preference have been noted,  
e.g. large hydrophobic residues in the P1 and P4 positions, and  
pro in the P2 position. Best known for its action on caseins,  
although it has been shown to hydrolyse hemoglobin and oxidized  
insulin B-chain.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
SUBTILASE FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X14130; CAA32350.1;  
DR PIR: S06997; S06997.  
DR HSSP: P00782; 2SRT.  
DR MEROPS: S08.019;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR000209; Peptidase\_S8.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR Pfam: PF02225; PA; 1.  
DR Pfam: PF00082; Subtilase\_S8; 3.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Hydrolyase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;  
KW Transmembrane.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217  
FT ACT\_SITE 281  
FT ACT\_SITE 620  
FT DOMAIN 1867 1872  
FT PROTEINS.  
FT

SQ SEQUENCE 1902 AA; 200139 MW; 4B8DB844D8CDF7 CRC64;  
Query Match 4.1%; Score 437.5; DB 1; Length 1902;  
Best Local Similarity: 21.2%; Pred. No. 2.4e-09;  
Matches 457; Conservative 201; Mismatches 728; Indels 773; Gaps 105;  
QY 370 RVEKGIKGLTITGGAGQSAALDHNIGVQNGDGLKVO-----LAETLT-SLKWVTEN 422  
DB 3 RKKKL-SILAGIVAGALAVLPEVGEIQAKAISQTKSSLIANYTAFAKQATDT 60  
QY 423 LPAKEKVTGKTRLTDTKIGFTNDNGIDSKPYLD----- 458  
DB 61 TAATTNQAIA-TQLAAKGIDY-NKLNKVQODIYDVIVQSAAPASENGTLRTDYSSA 118  
QY 459 -----KDTGIHAGGQKITLTNGVYDDAATG-----QLKKNQT 494  
DB 119 ETQOETNKVIAQAASVKAABOVYTOOTAG-----ESGYVYVNGFTKVRVVDIPKLOI 172  
QY 495 AASALOTFTVKV-----DKNGDANDSKITVGNKPKDGTQVNTLKLKGENGVDT--- 547  
DB 173 A-GVKYVTLAKVYPTPAKANSMAVQAVSNKYKGEFTVSVT---DSGIDPTHKD 226  
QY 548 -----TEP-----NGVTFGLN-QNNGLVGNSLTNND-GLSVK 579  
DB 227 MRLSDDKVYKLTLSVKEFTPAKGRFENSKVYGEVYANNNDITDDYDDEQGHMVA 286  
QY 580 NTNSKQIQVAGD-----ITFDISNSKGA---GIENNT 612  
DB 287 GI-----IGANGDPPAKSVGVAPAEOLAMKVFPTNSATGSAITVAIESA 339  
QY 613 RTTRD-GIGFANNNGSIDANKPRLTPTGINAGGKELNVQSAINPANGQQLPMNR-- 668  
DB 340 KIGADVLYNLSGDSGNQDLEDELAANO-NANESGTAIVISAGSGTSGSATGVNDY 398  
QY 669 -----LSTANTEKSGSAATIKDLYNLSQVPLT-----FAG-- 698  
DB 359 YGLDQENMGVTPGTSRGATVYASAEVDVITQAVTTGTGLOLGPETQLQSDNDFGSF 458  
QY 699 -----DGPNTK-----KIGELIKYKGGTTADDTLK-----N 727  
DB 459 DQKRYVYVVDASGNLSKRVADYADAKGKIAVYRGELTPADOKQAQAAGALIVN 518  
QY 728 NIGV-----VADSDNLSLYKLAKTL----- 748  
DB 519 NQGTATPYTSNALTTFPTFGLSVYTGKLYDWAHAHDDSLGKIALITLVPNQKTEDEK 578  
QY 749 -SDLDVAV-TKTLTASDKRVYVDSGNNTAKLQNGDLTESKQNTGATP-ATNSKTIQVGLK 805  
DB 579 MSDFTSYGPAVSNLSFKPDITAPGGINWSTQNNNGYTNMSTGSMASPTIAGSQAALLKQALN 638  
QY 806 FTDNNGIA-----LQGTYYTKDKVGFAPAKQDGLSKRYLDKDLKLYGEVIT---NG 857  
DB 639 NKNKPFYVYKOLKGTAL-----TDELKT--VEMNTAOPIND 673  
QY 858 INAGKATITGLSNTLTDTATNATGHTVQGLVDSIDTKRAISIDVYNAGFNELKNNG--- 914  
DB 674 IN-----YNNVYSPRRGAGLVYVKAIDALEKNPSTVVAE-----NKYPA 715  
QY 915 -DAKDVSTYDTPVFINGNATTAQVY-----DGA-----SKVAVYVNDGTTI 958  
DB 716 VELKDFSTDKTFLFTFNRTTHETLYQMSDNTDTNAVYTSATDPNSGVLYDKKIDGAAL 775  
QY 959 -----HLTGADG----- 965  
DB 776 KAGSNITVPAGKTAQIEFTLSLPKSPDOQDFVEGFLNFKGDSGRMLNPLYWGFEGDNDG 835  
QY 966 -----NKNQIGVTKTTLTKTPAK-----DKAIFSVNSG 995  
DB 836 KIVDSLNITTSYPAGNGVYPLLTNNK-TGTQYVGGKAVYDADSNQYVDDQALIAF---SS 891  
QY 996 DDKALIN-----ARDIAD-----NUNTLAGEIRNKK----- 1021  
FT

Db 892 DNALYNDISMKYLLNINISNOVDILDGGANKVTLLSSSTNRKTTYNAHSGQYIYHA 951  
 QY 1022 -----GAPALOTFOYKKEKENGDD-----DNDADTT-VGKDA 1055  
 Db 952 PAMDTYYDORDGNIKITADGSGYRISGVPEGDKROVEDPFRKIDSKAPYRHAVALSA 1011  
 QY 1056 KTNQVNT---LKLKRG---NGLDIQ-----TKKDGTVFEGINTOSGLKAGNNTTL 1099  
 Db 1012 KTEGKTOYLLTAARAKDLGDLTKSVKTALENTVLLDTFTDAGTTADGY----- 1063  
 QY 1100 NNNGLSIKNTAGNEOIOVGADGVFAKVNNGVAGIDGTTTRITRDEIGFAGTNGSLDKS 1159  
 Db 1064 -----TKIETPLSDQQA-----LGNG-----DNSAEVLTPDNASATDQASVQ 1104  
 QY 1160 KPHLSKDG--INAGC--KTTINOSGELAIQNSNDATVIGKIYDLKLEKISSTAKTAQ 1215  
 Db 1105 KPGSTFPLIYNGGIGIPDKISSTTGYEANTQ-----GGGTYFFSGTYPAVVDGTYTDAQ 1159  
 QY 1216 NSLHEFSVA--DEGNNFNVSPSSYDTSKTSYDTFTFAGEGDI-----TTKVNKGV----- 1265  
 Db 1160 GKHDLLNTTDAATNFTASMPVTNADYAAQVLDYADKATHTQLKHDTKRLTAPFTD 1219  
 QY 1266 ---VHVIGIDOTKGLT-----TPRLVGNNGKVIDSON-----GONTIT 1303  
 Db 1220 LKFNNGSDQTEATIKYTGVSADTKTVNVDVA---ALDAQHFGVDVNVNGDWTIK 1276  
 QY 1304 GLSTPLANTNDKGSVRTEDGKTIKDEDKTRAASIVDYVSAGNLOGNEADVFTYD 1363  
 Db 1277 -----VATDEDCNT--TTEKRTITSSYD-----DMLK-----NSVTFQCGVTFG 1315  
 QY 1364 TVNFADGATTAKTAYTDDTSKTSKVYVDVNDPTTIEVKKKLGVK-----TTLTSTGT 1418  
 Db 1316 ANEF---NAISAKP-YDPKGTIATITGKVKAPHTTLLQDQKQIPKIDDLFFSTLIDLCTL 1371  
 QY 1419 GANKFAL-----SNOATGDALVASDIAVAILNLTSGDIQTAKASQANSAGYVDADGN 1472  
 Db 1372 GKPRGVVVGDTQNKTFQEGALFLLDAVA--PLTSLDSSV-----DAPVYTNPNPF 1421  
 QY 1473 KYIYDSTDN-KYQAKKNDGT-----VDKTEVAKDKLVAQAQTPDGLAOMNKSVIN 1555  
 Db 1422 QITGATDNAGYLSLSINGSSVASOYDININSGKPGHMAIDQ--PVKLLBKKNLVTV--- 1477  
 QY 1526 EOYNDANKKQIINEDNAPVKGLEKASDNKTKMAVGTGLANAAQPTFAGDTGTAK 1585  
 Db 1478 -AYVD-----SEDNTRYKNTIYYEPKTL-APPY-----TP-----STEP 1513  
 QY 1586 KLGETLITKGGOTDNTKLDNNIGVAGTDGFTVKLANDLTNLNSVNAAGTKIDD---KG 1642  
 Db 1514 AKTVTLTANSAT-----GELYQYASD-----GKTYQDVPAG 1547  
 QY 1643 VSFVDSOAKANPVLASANGLDLGGKVISVNGKTKTDTAANQOL-NEVRNLGLGNA 1701  
 Db 1548 VT-VTANGTFKFSYDLGNSPAAVYVNTI---KADDAQALQAAKQELTNTL--- 1596  
 QY 1702 GMDNADGQVNTADIKKOPNSSNRTVITAGVVLGKGNNDPKLKTG--GVQVGVK 1759  
 Db 1597 -----IASAKTISASGKYDATT-----TALAAA-----YQKATALDQIRNASVDS 1637  
 QY 1760 DGNANGDLSNVWKTQXGSKKALLATYNAAGQTYLTLNPAEALDRINEGIRFEHVND 1819  
 Db 1638 LTGANRDLQT-----AIHQLAAKLPADKKTSL--NQOQSKALELIDL----- 1679  
 QY 1820 GNOEPVVGRCNIDSSAGSKHVAIGFOAKADGEAAVAIGKTOAGNOSIAIGNAOKATG 1879  
 Db 1680 GNO-----TDSSTGK-----TFTALDDIVA-----QAQGTQT--DDQLQAT- 1715  
 QY 1880 DQSIAGTGNVAVKHSAGIGDPSTVKADNSYSVGNNOQEDATQTDVEGVGNNTTVES 1939  
 Db 1716 ---LAKVLDVAVLAKLAEG-----IKATTPAEVGNAK---DAATGKTYADIADITLISG 1762  
 QY 1940 NSVALGSN-----SAISAGTHAGTQAKSDGTAGTTTACATGT---VKFAG 1984  
 Db 1763 QASASASDKLAHLQALQSLKTVAAVAEAKTVGKGDTGTGSDKGGCGGTAPFPBGIG 1822

QY 1985 QTAVGAVSAGASAEERRIONVAAGEVSATST---DAVNGSQLKAKOGIANATNELDHR 2040  
 Db 1823 KDKGDEGSSPSSG-----NIPNPAITTTSTSTDTTDRNGQLTSGKALPRTGTTTER 1876

RESULT 13  
 OMPB\_RICRI  
 ID OMPB\_RICRI STANDARD; PRT: 1654 AA.  
 AC 053047;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)  
 DE (rOmpB) [contains: 120 kDa surface-exposed protein (surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiinae; Rickettsia.  
 OX NCBI\_TaxID=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R.  
 RX MEDLINE=92167802; PubMed=1724278;  
 RA Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;  
 RT "The 120 kDa outer membrane protein (rOmp B) of Rickettsia rickettsii is encoded by an unusually long open reading frame:  
 RT evidence for protein processing from a large precursor."  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RN [2]  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN=R.  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kDa surface-exposed protein of Rickettsia rickettsii."  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC - FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC - STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC - VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC - LAYER WITH HEXAGONAL SYMMETRY.  
 CC - SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X16353; CA34403.1;  
 DR InterPro: IPR003858; ROMP\_rOmpB.  
 DR Pfam: PF02708; ROMP\_rOmpB: 1.  
 KW Antigen, S-layer, Cell wall.  
 FT CHAIN 1 1333  
 FT DOMAIN 1334 1654 32 KDA BETA PEPTIDE.  
 FT CHAIN 1181 1188 POLY-THR.  
 SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 4.18; Score 435.5; DB 1; Length 1654;  
 Best local similarity 20.68; Pred. No. 2,4e-09;  
 Matches 376; Conservative 184; Mismatches 600; Indels 663; Gaps 86;

QY 630 ANKRLPTGNNAGKELTNQASINPATNGGQDLPNNRSLSTANTEKSGSAATIKDL-VN 688  
 Db 2 AOKNPLFKLLISAG---LVTAATAIYASFRGSA--MGAALQONRTTNGATVVDAGFD 56  
 QY 669 LSOVPLTFAGDTPGNVTKLGEILKVGKRTTADDLTKNNIGVAVDSTNSLTVLAKTL 748

```

Db 57 QTRAP-----ANGVALNANI-----IANANGINFTPAGSEFGL----- 93
Qy 749 SDDAVNTTLTASDKYT-----VDSNNMTAKLONGLTFESKONTGATPATNSKTTIGV 801
Db 94 --LNTANNLAVVSEDTTIGFTTNNVHNAHSEFLTLNACKTTLITGOGVTNNQAATKNA 151
Qy 802 DGLKFTDNGIALDGTFTYTKKVKFADKDSL--DKSPYLDKRLKAYGEVEITTINGIN 859
Db 152 QNVYVFPNNGAIDNNDLKGVIDGAPASTLVENLAPTTOKAPLILGDAVLANGVN 211
Qy 860 AGCAATLGSNTLTDATNATGTHVTLGIIVSTDKTRAASIDVYNAGFNLKNNGAKDF 919
Db 212 ----GLTANTNGFIQVSNKSPATVAKAINIADGO-----GIIFNTDAN----- 249
Qy 920 VSTYVDIVDEINGNATTAKATYTGKASKAAYDVAVNDGTTIHLTGADGNKNOIGVKTTLTK 979
Db 250 ----NANP-----LNLAGGTTTINFTGDTGTC-----RLVLSK 279
Qy 980 TPAKGDKAINFVN--SGDDKALINAKDIADNLNTLA--GEIRTKGTADTALQTPQV 1034
Db 280 HAA--ATNFTITGSLGNLKGVI-----EFNTVAVDGOILTANAGANAVI----- 322
Qy 1035 KVKENGDDNDADDTITVGDAKTNOVNTLKLKGNGLDIOTNKGDTVTGFIQTOSGLKAG 1094
Db 323 ----KTNNGAGRAAG 333
Qy 1095 NMTTLNNGSLIKNTAGNEQIOVGADVKFAK-----VN-NGVYAGIDGTTT 1141
Db 334 FVYSDVNGKVA--IDGOVYAKDWIQAANATGOVNFRIHIVDVGAGTT- 380
Qy 1142 ITRDEIGFAGTSGLSKSPHLSKDGINAGKKITINOSGETAONSDNAVTKGTYDLKT 1201
Db 381 ----AFKTAASKVT-----ITDQSNFGNDFGLMAQI 409
Qy 1202 ELEKNISSAKTAONSLEHFSVADQGNFT--VSNPYSSYDTSKTSVITF----- 1251
Db 410 KVPNAITLV-----GNFTGDA SNP-----GNVAGVITFDANGTLES 445
Qy 1252 ----AGENGITTKVKKGVVGVGIDQTKITPKLVGNNN-----GKGVIDSQNG 1299
Db 446 ASADANAVATNNTITAEISG--AGVVOISGTHAELRLGNAGSIFKLADGVINGKXNO 502
Qy 1300 NTITGLSMTLANVNDKGSVRTTEOGK-----IIRDEKTRAASIVDVSAG-- 1346
Db 503 TALVGGALAAAGTITLDGSAITITGIGNAGGAALORITLADAKTTLILGANNIGAGGG 562
Qy 1347 -FNLOGNEAVDFVSTYD--IVNFADGNATTAKYTYDDTSKTSKYVVDVNDTTEVKD 1403
Db 563 TIDLOANGGTIKLSTONNIYVDEPLATIDOTGVDAASLT-----NAQTLTINGKI 615
Qy 1404 KRLGVKTTTLTSTGAKKAPLSN-QATGDALVYKASDIVALHNLSTGDIQTKAKA--SQ 1459
Db 616 GTIGANKTKTLOFNIGSSKTYLSNCAVAINELVIGND-----GAVQAPADHTYLLTR 666
Qy 1460 ANSSAGYVADAGNKVIYDSTDNKYYQAKNDGTVDKTEVAKDLIVAQAQTGDTLQAMNV 1519
Db 667 TTNAAGQ--GKIIFN-----PVVYNGTJ--LAAGTINLGSATNP--LAELNF 706
Qy 1520 KSVIRKEOVNANKKOGINEDNAF--VKGLEKASDNKTKAAV--TVGDINAAVQOT 1572
Db 707 GS-----KGVNVDTVLNVGCVNLVATNITTITDANGSEFVFNAGTINIVSGT 753
Qy 1573 P-----LTFAGDTGTJAKKLG-----ETLIRKGGOTDNKLTNNNGVY 1611
Db 754 VGGOGGNKFNVALENGTIVFVLGNATFNGTITTAANSTLIDIGENVYADCAVADGTGIV 813
Qy 1612 --ACTGFTVYKLANDLNLNSV-----NAG--GTRKID----- 1639
Db 814 EFVMTGPITVTLNKOAPVNLKQITVSGPGNVVINEIGNGNHGAATVDITIAFENSIG 873
Qy 1640 -----DKGVSVVDSG-----QAKANTPYLSANGLDGSKYISVNGKGTKTDAANVOQLN 1690

```

```

Db 874 AVFELPRCIPENDAGNTMPLTIKSTVGKNTAKGPDVBSVVLVY-----DSVIADGVIG 928
Qy 1691 EYRNLILGNGNNDNA-----DGQOVNIADIKKDPNKGSSSNRTYIAGT 1735
Db 929 DONNVIGLG-LGSDNGILVNAVTLIYAGISLTNNNGVITLSCGVPNRP--TVYDLGT 983
Qy 1736 VLGR-----GN-----NDTEKLAGVOGVYDKG-----NANGD 1766
Db 984 GIGASKFQVFTTDDYNNLGNINATNATINDGVITVTTGI-AGIGFGKITLLGSVNGN 1042
Qy 1767 ----LSNVVVKOKGSKKALLATYNAAGCTNNYLNPNPAELIDRINNOGIRFEFVND 1819
Db 1043 VRFADGILSN--STSMGTITKA-----NNGVITYLGN-----AFVGN 1078
Qy 1820 GNOE-PVVOGR-NGIDSSAS-----GKHSVAI-----GEOAKAD-- 1851
Db 1879 GDSPTPVASVRFSTDSGAGLOGNITSGVIDFTGYNLGIYNSNITLGGTTALNGKIDLV 1138
Qy 1852 ----GEAAVAIGRO--TQAGNOSIAIGDQAQ- 1877
Db 1139 TMTLTFASGTSTWGNNTSIETTLTANGNIGHIVILEGAGVNTTGTITIKYODNANAN 1198
Qy 1878 -TGDSIAIGTGNVYVACKHSGAIGDP-----STVRADN--SYSGVNNQFETDA 1922
Db 1199 FSGTOTYTLIG--GARFNGLTSPNFAVYTGSRNRPVYSLIRANODIYITRTNAENV 1255
Qy 1923 TOTDVF-----GVGNNT--VTESNSVA--LGSNSAISA----- 1952
Db 1256 VTNDIANSFPGAGVGOVQNTVFVNATNTAAYNNLLAKKSANSANFVGIATDTSALT 1315
Qy 1953 ----GTHAGQAKKSDGTACTTTTACATGTGVKFAQOTAVGANSVAGSQA 1998
Db 1316 NVOLDLAKDIOAOLGNRLG--ALKYLGTPETAENAG-----PENAGISAAPAAG 1362
Qy 1999 ERRIONVAG-EVSATSTDA-----VNSOLYKA--TQIANATNELDRHONENKANAGI 2052
Db 1363 DEADINVAAYGIMAKPFYTDHOSKKGGLAGYKAKTGVVIGDPL-----ANONLM--I 1414
Qy 2053 SSAMAMAMP--QATIPGSMVTG-----IATHNGGAVAVGLSKLSDNGM 2098
Db 1415 GMAIGITKTDIKHODYKKGKTDTVNGFSLYGAQOLVKNFPAQGSALPSLQNVKKRSOR 1474
Qy 2099 VFKINGSADTQGHGAVGAGFH 2121
Db 1475 YF-----FDANGMNSKOIAAGHY 1492

```

## RESULT 14

```

P2P_LACPA      STANDARD;      PRT; 1902 AA.
ID 002470:
AC 002470:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (cell wall-
associated serine proteinase) (Lp151).
GN P2P.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_Taxid=1597;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing, and expression of the gene encoding the cell-
envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151."
RL T. Gen. Microbiol. 138:1353-1364(1992).
RN [2].
RP SEQUENCE OF 189-196.
RX MEDLINE=92226694; PubMed=1564442;

```



RA Naes H., Nissen-Meyer J.;  
RT "Purification and N-terminal amino acid sequence determination of the  
RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.  
RL *paracasei*."  
CC J. Gen. Microbiol. 138:313-318 (1992).  
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC GROWTH OF THE BACTERIA ON MILK.  
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
CC specificity, although some substrate preference have been noted,  
CC e.g. large hydrophobic residues in the P1 and P4 positions, and  
CC pro in the P2 position. Best known for its action on caseins,  
CC although it has been shown to hydrolyse hemoglobin and oxidized  
CC insulin B-chain.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
CC SUBTILASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M83946; AAA25248.1; -  
DR PIR: B44858; B44858.  
DR HSP: P00782; I501.  
DR MEROPS: S08.019; -  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR002029; Peptidase\_S8.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR Pfam: PF02225; PA; 1.  
DR Pfam: PF00082; Peptidase\_S8; 3.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; FALSE\_NEG.  
DR HydroLase: Serine protease; Cell wall; Zymogen; Signal;  
KW Transmembrane.  
KW Transmembrane.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902  
FT DOMAIN 188 1876  
FT TRANSMEM 1877 1895  
FT DOMAIN 1896 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT DOMAIN 1867 1872  
FT SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;  
Query Match 4.0%; Score 433.5; Db 1; Length 1902;  
Best Local Similarity 21.9%; Pred. No. 3; e-09;  
Matches 471; Conservative 231; Mismatches 815; Indels 631; Gaps 114;

Db 326 TGSATLVAISAEISAKIGADVLNMSLGSDSGNQTEDEPEIAAVOVAN----- 371  
Qy 305 KATKKNATDVFSIGNGNNGNSIRKTIINAGASADDVAVNAQLEAVRLANRLTFKG 364  
Db 372 -----ESGTAAYISAGNS-----GTSGSAQGVN-----KDYGLQDENW--G 408  
Qy 365 DSNRNREKGIKLLITGGAQTSALTDHNGVUNGDLKVOLEAETLSIKWTEENLT 424  
Db 409 TPGISR-----GAT-TYASANTDIVSQ-----ATIIDGKDLQGPETIGLS----- 450  
Qy 425 ANEVYVTKRLTDTKIGFTNDMNG-IDESKPYLDKDTGTHAGQKLTITLAVVDDAA 483  
Db 451 -----SNDFTSGSPDKKFFVVKD-----ASGD-----LSKGA-----AA 479  
Qy 484 TYGQAKKVNQTAESALQTFYKVKDKNGND-----ANDSKITTVGKNKPPDQTVNLT 536  
Db 480 DY-----TADAKGKIAIKRGLNPNADKQYAQAAGAGLITV--NNDGATPLTSTI 529  
Qy 537 KIKGENGVDTTEINGVTFEGLNONG--LTVGNSTLNGLSYKNT-----NSNKQIYV 589  
Db 530 RL-----TTFEFTPLSSKTGQKLVDTVAHPDDSLGKIALTLIPNQYEDK 578  
Qy 590 GADGTFPTDISN-----SKPGAGIENTRTITDGIQFANNQSLDANKPRLTPTGINA 642  
Db 579 MSDFSTGYSPVNSLSEKPDITAPGNIWS-----TONNNGYTMSTSMASP-----FIA 627  
Qy 643 GKELTIVQSAINPATNGQGLDPANRLSTANTE-----KSSGATITDL--YNLSQVLTIF 696  
Db 628 GSOAL--LKOALNNKNNFPYADYKQKGTALTDLEKVTYEMTAQPIINDINYN----- 677  
Qy 697 AGDTGPNV-----TKKLEILKVGKGTADDTLKNIGVA-----DSTNSLTV 742  
Db 678 -----NVIYSPRQAGLVYD-----KAALDLKRNPSIVAVENGYPVELKDTFTSTKTF 728  
Qy 743 KLAKT-----LSDLDVAVNTKTLASDKVTVDSGNNTAKLQNG--DLTFSSKONT 788  
Db 729 KLTFFNRTTHTLTYQMDNTPDINAATVPNSGVLTKIDGAIKAGSDITVPAGKT 788  
Qy 789 GATPPTNSKTGIVGDLKFTDNNGLALDGTYYITKDKVFAFADQSDLSK--PYLD--- 842  
Db 789 AQIEFTLSPKSFQDQQVE-----GFLNFKGS--DSSRLNLPYMGFE 830  
Qy 843 --KDKLXGEVEITNGIN--AGGKAITGLSNLTDAATNTATGIVTGLIVDSTDKTRA 897  
Db 831 DMND--GKIYDSLNGITTPSAGN-----YGIYPLITKNNTGHQYGVAV--TDADGK 879  
Qy 838 ASIGVLNAGFNLKNNGDPAKDFVSTYDVFINGNATTAKY--TYDGKASYAVYVNDGT 956  
Db 880 QTVDDQALN--ESSDKNALYNDISMQY--YLNRISMVQYDILDGQNKV-----T 926  
Qy 957 THTLTGAGNKNQIOYKTTTLTKT--DAKGDKAINFVSNSGDKRALINAKDIALDNLTLA 1014  
Db 927 TLS-----SSNTKRTIYDAHSOKYIYVNPADWG--TYVDOR----- 963  
Qy 1015 GEIRNTKGTADTALQTFQVKKVKNENGDD-----DNDADFTT--VGKDAKTQVNT- 1062  
Db 964 GNKK-----TADGSTTIRISGVPEGGDKRQVDFVFKLDSAPYRHALAKTENGTQ 1019  
Qy 1063 --LKLKGG--NGLDIO-----TNKDGTVTFEGINTOSGLKAGNNTLNNNGLSIK 1107  
Db 1020 YLLTAEKNDLSGLDATSVKTAINEVTLNDAITFDAGTADGY-----TKIE 1067  
Qy 1108 NTAGNQIOVADGVKFAKVNNGVYAGAGIDGTRITRREIGFAGNGSLDSKPHLSDG 1167  
Db 1068 TPLSDSOAOA-----LGN-----DNSAEIYLLDNASNAVNDASVQKPESTISFD 1112  
Qy 1168 --INAG--KRTITNOSGEIAONSNDVATGKIYDLKTELEKIKSTAKTAONSHEFSV 1223  
Db 1113 LTVNGGIPDKLISSTTTGEANTQ-----GGGTTFSGTYPRAVVGTYTDAQKKHDLNT 1167  
Qy 1224 A-DEQGNFTVSNPSSYDTSKTSQVTFAGENGIT--TRKVNKV-----VRGID 1271

Db 1168 TYDATNSFTASMAVFNADYAAQVLDYADRAHNOILLKHPDKVRLTAPFTDLKFNNGSD 1227  
 QY 1272 QTKGLT-----TPKLVGNNGKGLVISON-----GQNTITGLSNTLAN 1311  
 Db 1228 QTSATITKVGIVSSDRTKTVAVGDTVA---ALDAQHHSVDVFNVDNTIK-----VT 1278  
 QY 1312 VTNDKSGVRTTEQOKIIRKDEKTRAAISIVDLVSAGFNLOQNGEAVDFSVYDT-VNFA-- 1368  
 Db 1279 ATDEGNT-TTEQKITSSTDP-----DYLK-----NAV-----TFDGVAFGAN 1317  
 QY 1369 DGNATTAQVYDDTSKTSKVYVYVNDVTIEYKDKLGVK-----TTTITSTGTGANKF 1423  
 Db 1318 EFNATSAKF-YDPKVTATTTGKVKHPPTLQVDGKOISIKNDLTSFTSLDGLTGLQKPF 1376  
 QY 1424 AL-----SNOATGDALVKASDIYAHNLTSGLDITAKASQANSAGYDADGNKYVD 1477  
 Db 1377 GVVYGGDTQKKTQREALTFLDLDAVA--PTSLDSST-----DAVYNDNFQITGT 1426  
 QY 1478 STDNRKYVQAKNDGTVDKTKEVAKDKLVAQAOTPDGTLAQNVKSVINKEQVNDANKR--Q 1535  
 Db 1427 ATDNAQYLS-----LAINGSHVSAQYADININSKPGHMAIDQPVKLE 1470  
 QY 1536 GINDNMFVAGLEKASDNTKNAAYVGLNVAQTPLETFAGDTGTAKKLGTELTIG 1595  
 Db 1471 GKN-----VLTVAVTDSNNTTTKTIV-----YIEPKTLAAPTYPSP-----TTEP 1513  
 QY 1596 GQDTNKLTPDNNIGVAVGTGFTVKLAKDLTINLSVNAAGTKIDD--KGVSPVSSGQA 1652  
 Db 1514 AKYTV-LTAN-----AAAGEVVOYASD-----GGKTQVDPAGAVT-VTANGTF 1556  
 QY 1653 KANTPVLSANGDLGKGVISNVGKTKVDANVQOLNEVRNLLGLGNAGNDNADNGQVYV 1712  
 Db 1557 KFKSTDLVGNESPAVDYVYVNI-----KADPPAOLQTAQAKALTNL----- 1596  
 QY 1713 IADIKKDPNSSSSNRTVYIAGTVLGKGNDEKLAG--GVQGVADKDNANEDLSNV 1770  
 Db 1597 ISAKTILSAGKTDAT-----TALAA-----TQKQALDQTDASVSLTGTNRQLQ-- 1647  
 QY 1771 WVKTKDGSKKALLATYNAAGQNTYLTNPFEAIDRINEGIRFFHVNDGNOEPVVOGNN 1830  
 Db 1648 -----AINQLAAPKLPADKTSLL-----NQOQSKAL 1675  
 QY 1831 GIDSSASGRKSVAIQFQAKADGEAANAIGHQDQAGNOSTAIGDMAQTDQSIATGTNV 1890  
 Db 1676 GTPLGQDTPSTGKTFTALDLDVA-----QAGAGTQ--ADQLQA-----SLAKVLDAV 1723  
 QY 1891 VACKHSGAIDPSTVAKADNSYSVGNNOPTDANQTVFVGNNIITYTESNVALGSN--- 1947  
 Db 1724 LAKLAE-----IKAAPAEVGNK---DAATGKTWADADLTLSQASDASDKLA 1773  
 QY 1948 -----SAISAGTHAGTOAKKSDGATGTTTAGATGTVGAFAGOTVAGAVSVG---A 1995  
 Db 1774 HLOALQSLKTKVAVAEAAKTAGKDDTGTSDKGGGQGTAPAPBDITKDKGDDESQPS 1833  
 QY 1996 SGAEKRIQNVAGEVSAISR---DAVNSQLYKATQGINAMNELDHR 2040  
 Db 1834 SGG-----NTPKPAATTSSTDDTDRNGOHTSGKALPKTAETTER 1876  
 RESULT 15  
 OMPB.RICITY  
 AC P96989; STANDARD; PRT: 1645 AA.  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (TompB)  
 DE (Tomp B) (Contains: 120 kDa surface-exposed protein (Surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide)  
 OS OMPB OR SLF.  
 GN Rickettsia typhi.  
 OC Bacteri; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=94040787; PubMed=8224886;  
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;  
 RT Cloning and sequence analysis of the gene encoding the crystalline  
 RT surface layer protein of Rickettsia typhi.  
 RL Gene 133:129-133(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT Mapping of monoclonal antibody binding sites on CNR fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii.  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [3]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RX MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cieplik W., Peacock M.G.;  
 RT Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent  
 RT mutant deficient in processing.  
 RL Infect. Immun. 60:156-165(1992).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: L04661; AAB48987.1;  
 DR InterPro: IPR003858; rompa.rompb.  
 DR Pfam: PF02708; rompa.rompb.  
 KM Antigen; S-layer; transmembrane; Cell wall.  
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT FT 1354 1645 32 KDA BETA PEPTIDE.  
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).  
 FT CONFLICT 657 657 H -> N (IN REF. 2).  
 FT CONFLICT 842 842 V -> I (IN REF. 2).  
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).  
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).  
 SO SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;  
 Query Match 4.0%; Score 430; DB 1; Length 1645;  
 Best Local Similarity 20.7%; Pred. No. 3.8e-09;  
 Matches 420; Conservative 215; Mismatches 626; Indels 768; Gaps 101;  
 QY 227 LAVGLTATAKAASSIANGSNAQAIGFAATAVGSGTQVNLNGIA-----LGFSGVQLQK 280  
 Db 12 ISAGL-VTASATATVA-----GFGSVAMGAMVQNRRTMAATVVDAGF----- 55  
 QY 261 DNDVMAANVAVAPDDNPDNRKATFKKAGADVESIGNSNGDSIRKRIINVGAGSAD 340  
 Db 56 -DGTGAV-----NLPAVATNSVITANSNNATTFTPGNNLN-----SLFLDTAN 98  
 QY 341 TDAYNVAGLKAVALNARQITFKGDDSNRRKELGKTLITITGGAQTSALTDINIGVQN 400  
 Db 99 ILAVYINNTTLGHVTVN-VTKQGNFFFTT--GAGSKSLTT----- 136





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:28 ; Search time 88.56 Seconds  
(without alignments)  
4145.159 Million cell updates/sec

Title: US-09-813-214A-9  
Percent score: 10708  
Sequence: 1 NMHIKVIYFNKATGTFMAYV.....NCSADTQGHVCAVAGAFHF 2122

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL.19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1246.5	11.6	2059	16	Q9PD50
2	1185.5	11.1	2353	2	P71401
3	1039	9.7	2712	16	Q9F3X5
4	831	7.8	1953	16	Q98HJ2
5	730.5	6.8	1299	16	Q9F3X6
6	728	6.8	1190	16	Q9PC04
7	722	6.7	3930	16	Q98E20
8	681	6.4	1098	2	Q48152
9	670	6.3	1107	2	Q9F2D8
10	654.5	6.1	3705	2	Q9F285
11	651.5	6.1	2240	16	Q9ZDP1
12	622	5.8	6713	16	Q99054
13	619	5.8	6713	16	Q931R6
14	604	5.6	2065	2	Q93DC7
15	600.5	5.6	3659	16	Q98LN6
16	590.5	5.5	2271	16	Q99QY4

17	565	5.3	3241	2	Q9AHF9	Q9ahf9 fusobacteri
18	554	5.2	2106	2	Q9XC47	Q9xc47 rickettsia
19	549.5	5.1	2586	5	Q9YTK8	Q9ytk8 drosophila
20	548	5.1	4919	2	Q9ZHL0	Q9zhl0 haemophilus
21	546	5.1	1557	2	Q9RNI2	Q9rni2 haemophilus
22	539.5	5.0	2703	16	Q9K0T0	Q9k0t0 neisseria m
23	536.5	5.0	4152	2	Q9ZHL3	Q9zhl3 haemophilus
24	534.5	5.0	1291	16	Q9ZK07	Q9zk07 rhizobium m
25	533.5	5.0	2478	2	Q9LCH2	Q9lch2 staphylococ
26	530	4.9	3890	16	Q99U53	Q99u53 staphylococ
27	527.5	4.9	5627	16	Q9I120	Q9i120 pseudomonas
28	525.5	4.9	2478	2	Q9RL69	Q9rl69 staphylococ
29	525	4.9	2147	2	Q9L950	Q9l950 pseudomonas
30	525	4.9	2402	2	Q9AER7	Q9aer7 staphylococ
31	519.5	4.8	2468	16	Q9I2M3	Q9i2m3 pseudomonas
32	518	4.8	2747	2	Q9L800	Q9l800 aeromonas s
33	517	4.8	2481	16	Q99Q86	Q99q86 staphylococ
34	517	4.8	3029	16	Q95582	Q95582 synechocyst
35	513	4.8	2514	16	Q9JY30	Q9jy30 neisseria m
36	508	4.7	1536	2	Q48031	Q48031 haemophilus
37	508	4.7	2349	2	P94750	P94750 escherichia
38	508	4.7	2383	16	P76347	P76347 salmonella
39	497	4.6	1963	2	Q9XC03	Q9xc03 salmonella
40	492.5	4.6	1618	2	Q9KKB1	Q9kkb1 rickettsia
41	491.5	4.6	1477	2	Q48028	Q48028 haemophilus
42	488.5	4.6	1619	2	Q9KKB7	Q9kkb7 rickettsia
43	488.5	4.6	2035	2	Q9XCJ4	Q9xcj4 salmonella
44	487	4.5	2479	16	Q9A988	Q9a988 caulobacter
45	482.5	4.5	1604	2	Q9KK99	Q9kk99 rickettsia

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	2059	AA.
ID	Q9PD50				
AC	Q9PD50				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	SURFACE PROTEIN.				
GN	XF1529.				
OS	Xylella fastidiosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;				
OC	Xylella.				
OX	NCBI_TaxID=2371;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=9A5C;				
RX	MEDLINE=20365717; PubMed=10910347;				
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,				
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,				
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briston M.R.S.,				
RA	Bueno M.R.P., Colombo A.A., Camargo L.E.A., Carraro D.M., Carter H.,				
RA	Colautto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa C.M.,				
RA	Coulinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,				
RA	Faciundani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,				
RA	Fraga J.S., Franca S.C., Franco M.C., Frome M., Furlan J.A.,				
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,				
RA	Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,				
RA	Kriegler J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,				
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,				
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,				
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,				
RA	Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,				
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,				
RA	Nhani A. Jr., Nobrega F.G., Nunes S.A., Palmieri D.A., Paris A.,				
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,				
RA	Petxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,				
RA	Peugot R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,				
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,				
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,				

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubaho M.H.,  
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,  
RA Zaio M.A., Zaitz M., Meidanis J., Setubal J.C.,  
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
RL Nature 406:151-159(2000).  
RW EMBL, AE003582; AB84338.1; -.  
KW Complete proteome.  
SQ SEQUENCE 2059 AA; 204035 MW; BBA671B5DD24E10 CRC64;

Query Match 11.68; Score 1246.5; DB 16; Length 2059;  
Best Local Similarity 24.38; Pred. No. 5.8e-33;  
Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108

QY	62	NGSVAAGIGISEADGG-----KGGANNRGOKSTAI-----	GDIAQA	97
Db	36	NAQVYINDGM---DGGCORYIDNCSAGVERTVATQCSDEDPWTVTGYARFFPGSPGTAEQ		92
QY	98	LG-SQSIAIGDNKTVHNSNNNANIGAKASGNESIAIG-GDVL-----ASGHASIAIGSDIL		151
Db	93	OGASRNLTIGGSLYV-----NSGQYGVNDYLKRTYSIMMGVYITMNYTAIGNALTAIGS---		145
QY	152	YLKREYVOQISELLPIIRGOKALNDIYQIADPNLQKTRBTHAGHASTAVGA--MSYAKG		209
Db	146	-----AASSADADLAKSLATKASG		164
QY	210	HFSNAFGFRATDEGTSYIAVGLTATA-KAASIAVSGNMAOIGAIPAATVAGSGYVNLNRG		268
Db	165	ARALAIAGKASADVDYVVALIGSGATCTAGTASSIAIGINASAVN-GAAVAVGGAALVYIPDG		223
QY	269	-IALFGSOYLQKDNVDYNAANVRAYPADDNQPIDNRKATFKFGATVDFSIGNSGNDSI		327
Db	224	AVALGINSVA-----STGKLGSGYDP-KTKTSTDSAAAMKSTILAAY-SIGDVSSTNLK		275
QY	328	RRKIIINGAGSADPDYANVAQKREAVNLAR--QITFKGDDSNRRVE-----KGLKJTL		379
Db	276	TROISGLAAGTSINDAVNAQKRYDEIARSGMNLTAAGNSGNAVAGSSVDLKNDDKIL		335
QY	380	TIT-----GGAQTSALPDHNIQVYOND-----GAKYOLAET---LTSIKMYTTE		421
Db	336	TITFAIGSNDYQFNLKNDYKVTTLAVGDALLNTDGIALGTDVSLSTTGIAITDGPVATAS		335
QY	422	NLPANENV-----TVCKTRLTDTKIG-----FTDMMGIDES-----		453
Db	396	GIDAGSKVISHVAAVASET--STDVANGSOLNAVQOASOPVTEFTGNEGAVKRSIGQSV		453
QY	454	-----KPYLKDQTS-IHA-----GGOKITKLAVGYDD		481
Db	454	VIGESSTACTYSGGNLKSAYVDEAGRIHIQIADSPKFGNVYINNGKISGVATAGEYD		513
QY	482	AATYQGLKRVNOTAESALQTFYVKYVDKNGDANDSKITLIVGKNNKPDGTQVYTLKIGE		541
Db	514	AVNFSOLKSTIAVD---QGWTLTLAGSNG--SKVASG-----TYDLKNT		554
QY	542	NGVDYTERNG---TYTFGLNQN---NGLTVGNSTLANDGLSVKNTNSNKQIOVGADGIT		595
Db	555	DG-NLITSKSGSDNDVVENLSKPKYDGMSTGYVNNNGVKY-----GSDVALGTTGLT		608
QY	596	FTDISNSKPGAGIENTTRITRDGIGFANNTGSLDANKPRLTFPGINAGGELTNVOSAI-		654
Db	609	ITD-----GPVATIASGIDAGSKVISHVAAV		635
QY	655	-----NPATNGQLDFMNLSTANTEKSGSAATIKDLYNL.SQV-----PLTFAGDTGPNYK		706
Db	636	SETSTDAVNGSQL-----NAVVOVASOPVYFTNGEGA-VKR		670
QY	707	KIGELIKVAGKRTADDLKRNIQVAVADSDNSLTVYKLAKT-----		747
Db	671	SLGOSVAVIGESSTACTYSGGNLKSAYVDEAGRIHIQIADSPKFGNVYINNGKISGVTA		730
QY	748	-LSDLDVAVMTKL-----TASDK--VTVDGSGNNTRKL-----QNGDLPFFSKONTGA		790

Db	731I	GTREEDAVNEBQLKSI	STAVDQWMTLASG	SGSKVASG	GVYDLKNTDGLTISK	SGD	SN	790			
QY	791I	TEATN-SKI	IGYDGLKFTD	NNGIALD	GTITYTK--KVF	AKODGSLDKSP	LYDKOKIK	847			
Db	791I	DVFNLSKDFK	YDGM-----	TSGT	VYVNDGVVVG-----	SDVALGTTGLT	831				
QY	848	VEE-VEIT	NGINAGKAITG--	LSNLT	PDATNATGHV	OLGIV	STOKTRASIG	901			
Db	832	FIDGP	AVIASGIDASK	YISHVAGV	SEISTDAVNG-----	SQ	LAOVYQASOP	PTFFG 886			
QY	902	DVLN	GFNLK	NGDKKDFV	STYD	IVDELING	NATTAKYTDG	KASKVAYDVNDGTTIHLT 961			
Db	887	---	NEGAVKRS	IGOSV-----	VISG	ESSTAG-TSG	GMKSVDEAAG--RIHQL 930				
QY	962	GADG	KNNGIGVYTTLT	FKTDAKGDKALINE	SVNSDDKALL	NADINDNLNT	LAGETR-N-T 1022				
Db	931	LADSK--FG-----	---	NVYING-----	---	GKISGT 952					
QY	1021	KSTADT--AL	QTFYVKRYKENG	DDDDADDTT	YTGKDATNQ	VNTLKLKNG	IGDIONKDG 1079				
Db	953	AGTEET	DAVNSQLKSI-----	---	STAVDQWMTLAS	GSKVASG--G 993					
QY	1080	YTTE	INTOSGLK	GNNTTLN	NNLSIKTNAG	NEQLOVADG--YKFA	VYVNGVYAGACID 1137				
Db	994	TY-----	---	---	DLKNTD	GNLTISK	SGSDNVFNLSKDFK	DGMTS 1036			
QY	1138	GTT	TRTDEIFFA-----	GTNGSLDK	SKPHSLDKINAG	KRTTIO	SEIIONSDNATV 1192				
Db	1031	GTTV	VNDGVYGVSDVAL	GTGLTGLIANG	PAVITASGIDASK	YISHVA	AAAVSETIDAVN 1099				
QY	1193	GK	ITDLKTELE	ENKISSTAKT	AOONS	LHEFSVAD	EOGNETVSNPYSSYDTS	SVITFA 1252			
Db	1091	GSQ	LAOVYQASOP	PTFFG-----	NEGAVKRS	IGOSVYIS-----	GESTAGTYS 1135				
QY	1253	GENG	TTTVYVNGGYRV	GDGTQKLT	PEPKL--TYG	NNNGGIYID	SONGNTTIGSLNLA 1310				
Db	1156	GGN-LK	SVYDPAAGR	RIHQLA--DSK	FE	GVYVINGGK-----	ISGT----- 1175				
QY	1311	NVT	NDKSVRTTEG	KITIKED	KTRAASIV	DLASG	FNLOGEAVDFY	STYDITNF--A 1368			
Db	1176	---	AGTEET	DAV--NF	SOLKSI	STAVDQWMTLAS	GANSKVASG	YDLKNT 1223			
QY	1369	DGN	NATLAKVYTD	DOTSKIS	KVYVDVND-----	DTTIE	VNDKILGKYTTLT	ISTGIGA 1420			
Db	1224	DGN	LITISK-----	SGSD	NVFNLSKDFK	YDGTSGT	VYVND--GVYGS	DVALGTTG 1275			
QY	1421	NK	FALSNQAT	DALVYKAS	DIVAH	LNLTSGDIQ	TAKASQANSAG	VYADGKRYVYDST 1480			
Db	1276	LTI	ANGPAVITASG	IDASK	YISHV-----	AAVASET	STDAVNSQ	LAOVYQASO 1326			
QY	1481	NKY	QAKNDG	VYDKTR	EVADKLYA	QAQTPD	GLAOMNK	YSYINKEQ----- 1527			
Db	1327	PVTE	FG-NEGAVKRS--	LGO	SVYISG	BSSTAGYISG	GNLKS	VYADRAAGTIHLQLADSKP 1383			
QY	1528	---	VNDANK	KOQI-----	NENDNAF--YK	LEKASDK	KYTKNA	AVTVG-----DL 1566			
Db	1384	GNV	VYINGGK	ISGVT	ACTEET	DAVNSQLKSI	STAVDQWMTLAS	GANSKVASG	YDYL 1443		
QY	1567	N	AVAQ	PTL	FAGDTGTTAK	LCETLTK-----	GGQTD	NKLTIDNNIGVAG 1613			
Db	1444	KNT	DGNLTISK	SGSD	NVFNLSKDFK	YISSEDKRE	SYTGNTOLDK	DKDVKYSSVWLLD	LSNELVITS 1503		
QY	1614	TDGE	---	TYKLAK	DLNLNLS	VNAGGTCTID	DGVSPV	DSGQAKNTPVYLSAN	GLDIGK 1669		
Db	1504	HS	STSVYK	TILANGES	SVYKRYV	VVGDVY	NIDB--VYV	NDLGLISTIVGAS	LTSLTINGSH 1561		
QY	1670	VIS	NGKGT	KD	TD	DAVNOQL	NEVRNLLGLG-NAG	ND	MDNADGNY--NTAD	LTK--DPN--- 1721	
Db	1582	KIT	VN	TAGTEED	TA	VNSQLKSV	SEAVD	GWMTLT	TSANGANSK	VYVSGGVYDLKNTD	GNLAI 1621
QY	1722	SG	SSNFTV	---	KACT	YVLG	GKGNNDTE	KLATG	GYQVGVYDK	DGANG---DLSN 1770	
Db	1622	SK	SD	SD	SN	DVFNLSKDFK	YDEYVAG--NT	VNTD	GVKYG	SDVSLGAM	LFTANPSV 1676

QY 1771 WKTQKDSKALLATATYNAAGOTNYLTNNPA---EADIRINEQIFREHYNDGNOEPVQ 1827  
 Db 1677 TASGFNMGDK---VISHVAAGMADTDVAVNSQLKQAVOSTVATKATRYSTINDG---TQ 1729  
 QY 1828 GRNCIDSSAGSKSHVAIGFOAKADGEAVALIGROTA-GNOSTAIGDNOAOTGDOSTAIG 1886  
 Db 1730 GGNVDGAGATGSKAIAAGVCTQASGEGAAVSGSAASGKSTAIRNALIASADGVALG 1789  
 QY 1887 TG-----NVAGKHSK-----AICDPSTVADKSYSGNNNOOTDAT----- 1923  
 Db 1790 DGAKDGRGAESEYTKYSGVQNNNTVGTGVSVDGAKGETRSISNVADKEMADAVNLKOLD 1849  
 QY 1924 -----QT-----DVEGV--GNNITVYES---NSVALGSNSAISAGTHAQ 1959  
 Db 1650 AVAKSNLQTDMDHREINNIEDYFKITKGSASSVKMGVNAALIGNAAVS-GTESVAL 1908  
 QY 1960 AKKSDGAGTATTAGATGTVKGAGOTAVGAVSVASGAEERRIONVAGEVSATSTDVAV 2019  
 Db 1909 GKNTNVASADNAVAIG-NGSVADRA-----NSVSYGSGSEPOVTNVAAG---TADTDAVN 1959  
 QY 2020 GSQYKATGCIANAATNEDLRIHONENKANAGISSAMAMSPQATIPGSMVTGCIATH 2079  
 Db 1960 VSOL---NGLITAKQYTDGAVGNLRRETSGGVAAAATATNLPOAVYQOGRMTSVGVSSY 2016  
 QY 2080 NGGAVAVGLSKLSDNGOMVFKINGSADTQGHVGAAGAGHF 2122  
 Db 2017 QGQSAIAVGSVAVSESGHWYFKPSGANTRSHVGVGAGVGYOM 2059  
 RESULT 2  
 P71401  
 ID P71401 PRELIMINARY; PRT; 2353 AA.  
 AC P71401:  
 DT 01-FEB-1997 (TEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TEMBLrel. 02, last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)  
 DE HSF PROTEIN.  
 GN HSF.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 NC NCB1\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS4;  
 RX MEDLINE=97047989; PubMed=8892830;  
 RA St Geme J.W. III, Cutler D., Barenkamp S.J.;  
 RT "Characterisation of the genetic locus encoding Haemophilus influenzae  
 RT type b surface fibrils."  
 RT J. Bacteriol. 178:6281-6287(1996).  
 DR EMBL; U01852; AAC44560.1;  
 SQ SEQUENCE 2353 AA; 243845 MW; 53E491B51A897B77 CRC64;

Query Match 11.1%; Score 1185.5; DB 2; Length 2353;  
 Best Local Similarity 22.7%; Pred. No. 6.4e-31;  
 Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;

Db 120 -----STNASSF-----TYSLKDLDTLT 138  
 QY 235 AKASSIAVSNAGALGFATAVG-----GNOVNLNGLIALGFSOYLQKNDVNNAN 288  
 Db 139 SVATEKLSFGANDKQVDITSDANGKLAKTANGNHLN-----GLDSTLPDAVNTGVIS 193  
 QY 289 VRAVAPDDQPIIDNRKATFKNGADVFSIG-NSNGNDSIRKRIINVGASADTAVNVA 347  
 Db 194 SSSFETPD---VEKTRAAIVK---DVLNAGNNG-----AKTAGANV- 230  
 QY 348 QLEAVRL-ANBQITFKGDDSN---RVEKGLKTLTITGGAOTSAITDHNIGVONG 401  
 Db 231 ---ESVDLSAVNNVEFIIGDKNTLDVVLAKENGKTEFEVKFPKTS-----VIREK 279  
 QY 402 DGLKVLAEFLSLKVTENTLANKYVVGKTRLTDDIGTNNMNGIDESKPYLDDT 461  
 Db 280 DG-----KLETKENNDNKVT---SNATDN---TDEGNGLVTKAAVID--- 318  
 QY 462 GHAGGOKITKLAGVDDDAATYGQLKRVNOTAESALQTFYKKVKDKGN-----DA- 514  
 Db 319 AVNKAQMRKVTITANQONDPAIVA---SGTNYTFESGDDTASVTKDITNGNGITVYKDAK 376  
 QY 515 -----NDSKITVYGNKNNKPDGTQVNTL---KLKGENGVDTT----- 548  
 Db 377 VGDGLFEDSKKIVADTITALTITVGKVAIEAKEDDKKLVNAGDVLVTALGNLSMKAKAEA 436  
 QY 549 EENGITVTEGLNONGITLVGNSITLNDGLSVK---NTNSNKQIQVCAOSITTTDSNSKP 604  
 Db 437 DTDGAL-GLSKDQEVKAEYTFKAGKNLKYKQDANFTYSLQDALTGLTITITIGTTN 495  
 QY 605 GAGIENTRTTRPDIGIF--ANNNGSLDANKPRLTPTGINAGKELTNVQASINPATNGQ 662  
 Db 496 G-GNDAKYIANKDGLITITAGNGTITGTITISYTKDGIAGKAITNVASGLR-AYDDAN 553  
 QY 663 LDEMNRLST-----ANTE---KSSGAATIKDLYNLSQVPLTFAG- 698  
 Db 554 FVPLNNSAYDLNREHVDAYKGLLNLEKNANKOPLVDTASTAVGDLRKLGVAVSTKNGT 613  
 QY 699 DTGPNYTKKLGELIKKYGKGTADDLTKNNIGVADSTNSLTVKLAKLSL-----D 752  
 Db 614 KEESNOYKKADEVLFTGGAATVYTSKSENG-----KHTTVSAVETKADCGLEKQD 665  
 QY 753 AVNTKTLTA-SDKVTVDSCNTAKLQNGDLTFPSKONTGATPA----- 793  
 Db 666 TIKLKVNDNTDNLVLTGVNNGTRAVTKG---PEYVTGATDADRKGVYKADATANDADK 722  
 QY 794 -----TNSKTTIGVDIKFTDN---NGIALDGTYYTTRKRVFAKOD 831  
 Db 723 VATVKVATAINSAATFVTEENTLTSTIDEDNPTDNGKDALAGDILFFKAKNLKVKRD 782  
 QY 832 G---SLDKSFYLDK-----DKLVG-----EVELTNGIN-----AG 861  
 Db 783 GKNTFDLAKNLEVKTAKVSDTLTIGNTPGTGATPKNVITTSADGJNFAKETADASG 842  
 QY 862 GKAI---TGLSNLTLD-ATNATGHTVQLGIVDSQKTRASIGDVNAGENLKNNDAD 918  
 Db 843 SKNVYIKGIAITTLTFSAGAKSSHVL--NVDATKSNASIEDVLRAGMNIQGNMVD 900  
 QY 919 FVSTYDTVDPI----- 929  
 Db 901 YVATYDTVNTDSTGTITVYVQKADGKADVKIGAKTSVKDHNKGLFTGDKLDANN 960  
 QY 930 ----- 929  
 Db 961 GATVSEDDGKDTGTLVTKVTVDAVNSGMRVTEGATETGATAVNAGNAETVSGTS 1020  
 QY 930 ---NGNATTAKTAYDKASKVAAYDVNV-DGTTIHLTGADGKNQGVTTTLTKTDK- 983  
 Db 1021 VNEKNGNATTATVSKONGINVKYDVNVGDGLKI-----GDDKKIVADTTLTVTGKV 1074  
 QY 984 ---GDKAIFSVNSGDDALINAKDIADNLNTLAGEIR-----NFKGTAIDLQ--- 1029  
 Db 1075 SVPAAGANSVA-----NNKRIVAEGLATALNLSMTAKADYYAOGSESGEITDQGVKAGD 1128

```

QY 1030 --TFQV--KRYKEN-----1039
Db 1129 KTFKAGKMLKQSEKEDTYSLODTLTLGLTSLTGLTANGRNDGTYNKQGLITTLAN 1188
QY 1040 ---DDBDADTLTVGKD-----AKTNOV 1060
Db 1189 GAAGAGTASNGMTISVTYKQISAGNKEITNVKSALTKYKDTQNTADETDOKREHFAVKN 1248
QY 1061 NTLKLGKRGGLDIQTKKD---TYTFGINTOSGLKAGNNTILNNG---LSIKTNAGNE 1113
Db 1249 NVEFEFGKGAATVSAKTDNNGKHTYIDV---AEAKVGGLEKDTGKIKLKADNTDGN 1305
QY 1114 QIOVGADGVKPAKVNNGVGA-----GIDGTRIRTRDEIFAGTNGS---LDKSKPHL 1163
Db 1306 LFTV--DARKGASVAKGEFNAVTTDAITAGTANERGVVYKSGNATATETDK-----1358
QY 1164 SKDGINAGKKTITNIOGSEIAONSNDVATGKIYDLKTELENKISSITAKTQNSLHEFSV 1223
Db 1359 -----KQVATV--GDVAKAINDAATFVKV-----ENDSATI-----DDSP 1392
QY 1224 ADEQGNFVSNPYSSYDTSKIDVTF--AGEGKITTKVKNKV---VRGIDOTKGLTTP 1279
Db 1393 TDDGAN-----DALKAGDTLTLKAGKMLKVRGKNTTFALANDSVKSATVSD 1441
QY 1280 KTLVGNNGK-----GIYIDSONGONT---ITGLSNTLANVTNDKSVRTTEQOK 1326
Db 1442 KISLGTNGKKNVITSDTKLNFRKADSKTGDANDIHLNGIASLTLDLNSGAT--TNLGN 1500
QY 1327 IKDEKDTAASIVLVLSAGFLQG-----NGEAVDFSVSTYDTVNFADGNATTAKVT 1378
Db 1501 GITDNEKRAASVYKLVNAGMNVGVKPAASANNQVENIDFVATYDTHVDFSGDKDTSVT 1560
QY 1379 YD--DTSKTSKVVYVNDVDTIEVKDKLGVKTTTLTSTG---TGAN--KFLSQAT--1430
Db 1561 VESKDNKRTFV-----RIGATSVYIKDHNGKLFKGKELKDANNNGVTV 1604
QY 1431 -----GDALVKAADIYAHLLTSGDIQTAAGASQAN-----SSAGYVADVKNVL 1475
Db 1605 TETDGDENGLVTAKAVIDAVNKAQMRKTT--GANGQNDDFATVASGTVTADNGNT 1663
QY 1476 YDSTDNKYYQANKD--TVDKTEKVAK-----DKLVAQA---QTPDGLTQNNVKSVIN 1524
Db 1664 AETV-----KANDGSITVYKYNKVDGLKLDGDKIVADTVTLTVADGKVTAN-----1711
QY 1525 KQOVNANKKQGINEDNAVVKLEKASDNKTKNAAVTYGDLNAVQPLTFEGDTGTTA 1584
Db 1712 ---NGDGRK-----FVDASGLADALNKLSWTA--TAGKEGTGEVDPAANSAGQ---E 1754
QY 1585 KRLGELTITKGGOTDNTKLTNNIGVAGTDGFTVLAKDLTMLNSV-----NAG-----G 1635
Db 1755 VKAGDKVTFKAG-----DNLKIKQSGKDEYTSKLKELKDLTVEFRDANGGTSSES 1805
QY 1636 TKIDKGVSEFVDSGOA-----KANTPVLASGLDGVKVISVNGKGT-----DT 1681
Db 1806 TTKTKGLTIPANGAGAGANTAMTISYTKDQISAGNKAVTNVVGKLFKFGDHTLANG 1865
QY 1682 DAANYOOL--NEVRNLLGIGNAGNDN---ADGNQVNIAD-----IKKDPSSG---1724
Db 1866 TVADFEKHYDNMAYKDLJLNDKRGADNNPTVADNTAATVGDRLGLNVIISADKTTGEPNOE 1925
QY 1725 -----SSNRVYIKAGT-----VLGG-----KGNPDTEKLTAGG---Y 1753
Db 1926 YNAOVNANAEVFKSGNGINVSCKTLNGTRVITFELAKEGVYKSNFTYKNAQSGETNLV 1985
QY 1754 QVG-----VDKDGANGLDSNVVKTQK-----DGSKALLATYNAAGTNNL 1796
Db 1986 KYGDWYTSKEDIDPATSKPMTG--KTEKYKVENGVYKVSANGSKTFTVTLNKGSG---YV 2039
QY 1797 TYNPAEADIRINEOGIRFTHVNDGNQEPVVGNGIDSSSGHSAVIGQAK-----1849
Db 2040 TGN--QVADAIKSGFEL-----GLADAABEAKAFESAKDKQLSKDKA 2081

```

```

QY 1850 ---ADGEAAVAIGRQTOAGNOSIAIGDNAQATGDO-----STAIGTNVAKHSG 1897
Db 2082 EYVNAHDKVRFANGLTKV---SAATVESTDANGDKYTTTFVKTVDLELPTQY---ND 2135
QY 1898 AIDDPSTVAKADNSYVGNNOFTDQTQDVEGVGNNTYTYESVALGSNSAISAGTHAG 1957
Db 2136 ANGNKIVKKNADGW-----YELNADGTASKEVTLGVNDANGK-----2174
QY 1958 TOAKSDGTAGTTTGTAGCTGVKAGQOTPAVGAHSV-----ASSAERRIONV 2005
Db 2175 VYKVTENGADKRYTNADADAADTKGEVSNKYVSTDEKHYVRLDPNOSNGGVYIDNV 2233
QY 2006 AGEVASTSTDAVNGSOLYKATQIGIAN--ATNELDRHONENKANAGISSAMAMAMP 2062
Db 2234 ANEISATSTEDAINSGQLAVAKGVITNLAGQVNNLEGKVYKVRADAGTASALASQLP 2293
QY 2063 QAVTPGRSVYTGIAITHNGOGAVALGISKIDNGQWYFKTNGSADTPQGHGAAVGAFHR 2122
Db 2294 QATMPGRSMVAIAGSSYOGNGLAIGVSRISDNGKVYIIRLSTTNSQKTVGAAGVGYOW 2353

RESULT 3
Q9F3X5 PRELIMINARY; PRT: 2712 AA.
ID Q9F3X5;
AC 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE MAFB PROTEIN (HSF).
GN MAFB OR HSF-1 OR PM0714.
OC Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=PM70;
RC MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM0.",
DR Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277636; CAC14203.1; -.
DR EMBL; AE006108; AAK02798.1; -.
KW Complete proteome.
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FEA3D CRC64;

Query Match 9.7%; Score 1039; DB 16; Length 2712;
Best Local Similarity 23.6%; Pred. No. 4,3e-26;
Matches 650; Conservative 325; Mismatches 871; Indels 904; Gaps 136;

QY 11 KATGTEMAVAVYAKSHSTGGSCATGQVGSVFTLSFARIAALVIVIGATLNGSAVAGT 70
Db 221 EGTTTYTGI--KYRFVAKSTLDDAVALDQ--DSVAIGPKAKTEGVAVALAGHDSITMAIESIA 278
QY 71 ISEADGKGGANRGDKSIAIGDIAQALGSQSAIGDKNIHVHNSNNANIGAKASGNESTI 130
Db 279 I-----GKKAIVATANKGIALIGEHAR--VSGKQ-----DGDVYVYGASSYVLGPK--DDESSV 328
QY 131 AIGDVLASGHASIAIGSDDLYLKKTETVOQISELLPIIRGKALNDIYO--LADTN---LQ 186
Db 329 AIDDKAVSRSEASIAIGKNAITSNKDAKQIA-----KNNIILGTNAQAIASDNSIALG 382
QY 187 KYRRTAAGHASTAVGAMSTAKGHFSAFCTRAATBCTYSILAVGLATATAKAASIAVGSN 246
Db 383 NAATNKDSDSIAIGDSATETKAHSLAVGTTSKALAEALALAGKLAELAKGTSSVAMGNT 442

```





Db 2368 PRNTVDLNTDGNISVKTNTADKHNVT---FGLADNINVKDSVYVGPKGANGKEGEGAV 2424  
 QY 1857 AIGRQTOA-GNOSIAIGDNAQATGDSIAI-----GTGNV-----VAGKHSQA-IGDPS 1903  
 Db 2425 VINAEBCANGKKGISI---VGKDGKQAVAIISGKDKGVTGICLGPAGADCKNNANLIIIVND 2481  
 QY 1904 TYKADNISYVGNNOFTDATTGDFGVGNNTITVESNSVALGSNSAISAGHTAQAK-K 1962  
 Db 2482 SYVGLDGNCKDGNKSTRIVYTKPNCEEDOV-ATMNDGLVFADKTEHAKLGTTVVK 2540  
 QY 1963 SDGTACTTTAGATGTVKGFAGQTAAGAVSVGS-----GABRRION-----VAAGEVSA 2012  
 Db 2541 GDDKNIETEVAGDTTIVR-LKQNDIVKGINVTENLTAKGCAKINNNVINDVADGEVNA 2599  
 QY 2013 TSTDVANGSOLYKATOGI---ANATNELDHRHONENKANAGISSAMAMASPOAYTPGR 2069  
 Db 2600 TSMQAVANGSOLKHYQVQVNNQATFINKLDHINKYDKDLRAGIAGATVAFLQRPBEAGK 2659  
 QY 2070 SMVTGGIATHNGGAVAVGLSKLSDNGQWVEFKINGSADTQGHV--GAAYG 2117  
 Db 2660 STVSLGVGSYRSESAIVAGYARNSDNNKISIKLGGGMNSRGDVPNGSGIG 2709  
 RESULT 4  
 Q98HJ2 PRELIMINARY: PRT: 1953 AA.  
 AC Q98HJ2: MEDLINE-21082930: PubMed-11214968;  
 DT 01-OCT-2001 (Tremblrel. 18, Created)  
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
 DE MLL2848 PROTEIN.  
 GN MLL2848.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 NC NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MAF303099;  
 RX MEDLINE-21082930: PubMed-11214968;  
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matenabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RT DNA Res. 7:333-338(2000).  
 DR EMBL; AF003000; BAB49874.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;

Query Match 7.8%; Score 831; DB 16; Length 1953;  
 Best Local Similarity 22.5%; Pred. No. 1.7e-19;  
 Matches 540; Conservative 251; Mismatches 828; Indels 780; Gaps 107;

QY 17 MAAVEYA-----KSHSTGGSCATGOV-----GSVPLSPFARIALAVIIGAT 60  
 Db 39 MAYLRRAHRYLGRKRGSGIGMIGIRVLAALGAILGATPALAQYAA-----GGGT 93  
 QY 61 LNSGAVAGIGISEADG-----KGGANARGDKSIAIGDIAO-----ALGSQSTAIAGD 107  
 Db 94 ANSAGSVAVGPSTATNGLGIAGVSGATQAGIDISIAOGTSANAGSONAVAIQFQSTIASQL 153  
 QY 108 NKTIVHNSNNANIGAKASGNESTAIAGDVLASHASIAIGSDLLYIKKETVO-QISELLP 166  
 Db 154 NSIYLQSRITAAAGTATA--QSAIGITDVTASQVDAIAVGRSSVASAQSVALGLSAKAT 211  
 QY 167 IIRGKALNDIYQIADTN-LQKRYRTHAOGHASTAVGAMSVYKSGHNSNFGTRATEGTY 225  
 Db 212 GTGGAMALGGGTIASINSVALGVQASATIGGAGNALGITSVASGSGNSTAVGTSSSAAAGL 271

QY 226 SLAVGLTATAFAASSIIVAGSNAOIGFAATAVGSGTOVNLNRGIALGFGSQVLOKNDVN 285  
 Db 272 SPAGWSVAAGSDAVVALGKSANAGLNSTALGSTNASADFAL----- 317  
 QY 286 AANRAVAPDQNOIDLRNRYKATFKNATDVFESIGNSGNSIRKILINVGASADFDVN 345  
 Db 318 -----GNOAVSSGISVAAGSG-SQATG 339  
 QY 346 VAOLKEAVRLANROITKGDSDNNRVEKGLKTLTITGAQTSALTDHINIGVQNGDGLK 405  
 Db 340 VS-----ATALGNMMAA-----TAANATFALIGAT--AGVDGTA-----IGKAN----- 378  
 QY 406 VOLAETLSLNAVTTENTLITANEKTYVGTRLTTP-----KIGFTDMNG-----IDESKPYLD 458  
 Db 379 ---ASAODALIMGTSARASSAOALAVGINAVATGKAVSVISGSGTAVGADAVSIGDPSYA 435  
 QY 459 KDTGTHAGGQKITKLITAGVDDAATYGOLKKNQTAESALQFTTVKVKDKNGDANDSK 518  
 Db 436 SGTGAFITGGANNIANSQGTATATA-----NOAGA----- 466  
 QY 519 IITVKKNNP-----DGTQVNTLKGE-----NGVDVTEETNGVTFGLNQNGLT 565  
 Db 467 -VAIGNNNKAIQGSVALGNGSTAGAAGLAGNVALGNCA-TAAASSGDVALGSGSVTAVA 524  
 QY 566 VG--NSTLNDGLSVKATNSKNKQIOVGADGI--TFTDISNKPAGIENTTRITRDSIGF 621  
 Db 525 VGTPAVINGTTTYAFQGTPTSTVSGAPGAERLTIVA-----AG-----RIS----- 568  
 QY 632 ANNCSILANPRLTPGGINAGKELTNVQSAIPATNGOLDPMNLSTAFNEKSGSAA 681  
 Db 569 GSSIDAVNGSOLFATNOAVDAIGTLLNNIN-----GGCIKFH-----ANSLADSSA 617  
 QY 682 TIKDLYNLSQVPLTFAGDTPGPNVTKKLOELIKVKGKTTADDLTKNNIGV--VADSTONS 739  
 Db 618 TGTDSVAL-----GPTSTATASAI-AAGSANAGASAIIGTSVASALDAT 665  
 QY 740 LTVKIAKTLSDLDVAVNTKTLTASDKVYVDSGNNNTAKIONGDLTFSKO--NGATPATNSKT 738  
 Db 666 AMGFSLKASGQFSTAVGANANATALSSTAIQONLASGVQATALQKANNGA-----SEA 720  
 QY 799 TGVDLKFTDNNGJALDQTYTITKDKVGFQAKODSLDKSRPYLDKLGVEYITNGI 858  
 Db 721 LALGNSTAGNAGVALGSGSVTAVNG-----TPNAVINGTTIYAFQGI 764  
 QY 859 NAGKAITGLNTLTATNATTTGHTVQIOLGIVSDTKTPRAST---GVNLAGEFLKNNGD 915  
 Db 765 NPASTVSGIAGCAERTLTNLAAGRISG---STPDVANGSOLFATNOAVDAIGTIVNN-- 818  
 QY 916 AKDPYSTDYDFINGNATTTAKVYDDKASKVAVADVNDGTHILTGADGKNKNGIG-VKT 974  
 Db 819 ---ISTGGIKRFPFANSTLA-----DSSA-----TGTD--SVAIGPTST 852  
 QY 975 TLTFTDAKGDKAINEFSVSGDDKALINAKDIADNLN-TLAGEIRNTKGTADTALQFQV 1033  
 Db 853 ATASATAIAGSNA-----NAGSANAISIGTSYVALDATTAMGFSLKASGQFSTA----- 902  
 QY 1034 KKVKENGDDDDADDTITVGDAKTNQVNTLKGKNGIDIOFNKDGTVTEGINTOSGLKA 1093  
 Db 903 -----VGANANAFALSTAI--GONAL-----ASGVQATALQKQANASA 939  
 QY 1094 GNNNTLNNNGISIKNTAGNE-QIOVGADGVYFAVYNNGVYAGIDGTRITRDEIGACT 1152  
 Db 940 SDALALGAN-----STAGNADVALGSGSVYAVAV--GTPNAVINGTT-----YAFQGI 986  
 QY 1153 NGSILDKSRPHLSKQIDINAG--KKITNOSGEIIONSNDVATGKIDYDLTELE----- 1204  
 Db 987 NPA-----STYSIGAPGAERLTILNLAAGRIGSSTIDVANGSOLATNOAVAIQITTV 1038  
 QY 1205 NKIS-----STAKTQNSLHESVADDEQNNFT-- 1332  
 Db 1039 NNISTGCGIKFYFHNSTLADSSANGTDSVALGTPSTA-TASALAIAGSNANASAGNASAI 1097  
 QY 1233 -VSNPYSSTDT-----SKTS-DVITTFAGENG-----ITTVKNGVYRVGID-----QT 1273







RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003004; BAB51100.1;  
 KW Complete proteome.  
 SQ SEQUENCE 3930 AA; 380662 MW; 9ACD1ACAI85BF712 CRC64;

Query Match 6.7%; Score 722; DB 16; Length 3930;  
 Best Local Similarity 21.4%; Pred. No. 1,3e-15;  
 Matches 589; Conservative 302; Mismatches 966; Indels 894; Gaps 133;

QY 16 FMAVAEAKSHSTGGSCATG-----OVSVRLSFARIALALAVIGATLNGSAYA 67  
 DB 1115 FSAVGGGTLTVTSGNVAAGLGPDEKGIQIAGM-TIGAGVNFSSAVAVDS--GGVST 1171  
 QY 68 GIGISEADGK---GGANARDKSIAIGDIAQALGSQ-----IAGDNKLYHNS 114  
 DB 1172 GIALSNSTSGNVTGGVDFRISGTAISLDNAASGTSENGTKVDPTVGGGGFLVQNS 1231  
 QY 115 NNNANIGAKASGN-----ESIAIGD-----VLASGHASIAIGSDIX 152  
 DB 1232 AAFVAVGNLVTNNAGADVSLFNNVTGIGIGITNSGNGVYVGGSAIIV-SADIS 1290  
 QY 153 LKKEVVOQIELPIIRGKALADIYQADPTNLOKTRRHAGCHASTAVGAMSAYAKHPS 212  
 DB 1291 SSATAPGTAVKVDGITGGSVTFSGLTSTGT-----GTGVSNTAAGS-G 1335  
 QY 213 NAFGTRATAGCTSLAVGLTATA-----KAASSIAGVS--NAQAIQPAAT--AVAGSNO 262  
 DB 1336 VGRGA-VTVSGAANGNGISGNAGSVTFGTCTVNTLGSANANAGIHFSGTNADVLGTTN 1394  
 QY 263 VNINRG-----IALFGSOVLKNDVNAANAVAPADNOPIDNRYKATPKRN 310  
 DB 1395 VTFGAGANQTDGIDFGSSASATAGFLTITGTGDLTSRGIDLSSTGNTKT-----TFAR 1448  
 QY 311 GATDVESIGNSGNDISRRKIIVAG-----SADPTAVVAQLEKAVRLANROITP----- 362  
 DB 1449 GSS-----ITNVGVELSSGCTTATS-----ANNFTFGDGN 1481  
 QY 363 KGDDSNRRVEK-----GLGKTL-----TITGAQTSALTJDNHIGVONG----- 401  
 DB 1482 AGDGLSTISAAAGYVTMTIGDPTLGNVNFDFPFGTSAHLASAVGTINWVSQNGYI 1541  
 QY 402 ---DGLKVOLETTLTSLKAVTTEJLANE-----KATVGRKRLTDTK-----IGFTNDM 447  
 DB 1542 AAGTGD---LSAAVTTISVACADALAGTQTFAFVGTVDLSTPTLDSGQSIITGFGNNT 1597  
 QY 448 N-GIDESKPYLDKDTGIIHAGGOKITKLTAGVVDVDDAATGOLKKNVQTMESALQ-----TF 502  
 DB 1598 SITVSCTIQPVANQNLGASGCVITG-NEGCV-TGTGDFLHLGSGNQVNTAFNPSGAG 1655  
 QY 503 TYAKVDKNGNDANDSKITTVGNKKNKPDGTQVNTLKLGENGVDTTETNGYTF-----GLN 559  
 DB 1656 SVFNVDSAGGFNNAGCIYI-----EGTVSVVAAGSTAFVAGAD 1696  
 QY 560 QNNGGLVGN-----STLNDGLSVKN-----TNSN 584  
 DB 1697 SNLSTNNVNVNAGTLLDVNGGTGNTTLRGLPNSAPPGTLTGGSISANSSGGLVNET 1756  
 QY 585 KOIOVGADG-----ITFTDISNSKPGAGIENTVTRITRDGIGFAN-NTGSLDA-NK 632  
 DB 1757 DKVTGCAVLSLGNSTGVTADDDITTSGA-----TAFSASGGGYVNTTGTINATNA 1811  
 QY 633 PRLTPTGIAAGKEJTNVOSAINPATNGQIDPMNRLSTAN-----TEKSGSAATTKD 685  
 DB 1812 QAAALDGIITAG---INFASSTAFASGNGIDLQNLSTASFGCTGLNTNGSGTGFNNGS 1867  
 QY 686 LVNLS--QVPLPFAFGDTGNTKKGLEILLKVGKGT-----ADDLTKNNIGVAVDSTDN 738

DB 1868 AYNLSGNAVISTGCTIASNGTAIVSIOELTGGSVTLSCGNTDGLAGAGNIVTGTIDN 1927  
 QY 739 ---SLVYKAKLTISDDA-----VNTK 757  
 DB 1928 GTAATVTFSGNRKQIDSGATDGVSLGNPNGLAFISGGLVITTSAGGAFHASTFGCTV 1987  
 QY 758 TLT---ASDKTVDSGNNPTAKLON-----GLTF---SKONTGATPATNSKTIGVDGLK 805  
 DB 1988 TVYKDGAGNNTITTTTGT-NALNIDNVTVGVGINFDSISSNCTGTGIALNNVSGAINIG 2046  
 QY 806 FPDNNGI-----ALDGTITIKDKYGF-AKDDGSLDKSPYLDMDKLYGVEVET- 854  
 DB 2047 AYDLBETIRGVDSVSGTGLSTLNTSLNGLAANAIGDLNLSAALGVSNITAGDVEDG 2106  
 QY 855 ---TNGINAGKAITG---LSNTLTATNATNTGHTVQLGI-VSDTQTRAASI----- 900  
 DB 2107 GSFAGTIGIDMAGTGTGTIGDPTVNNNPAGQSTIANVGVGFSSATNAQLVFGDGA 2166  
 QY 901 -----GDVLAGFNKKNNDADKDFVSTYDTVFT-NGNATTAKVTV----- 940  
 DB 2167 GPAESSIATGGQVJHADDLTPTSGD---YNFNDVNEGSDTSLSAVRYVYTAGGTG 2221  
 QY 941 DGRASK-----VAYDVNDGT--TILHGA-----DG----- 965  
 DB 2222 DGSILANPGSYLGAQASTANVYVLLDKNVGAQETITDLSGTFPNDLDDGVYLLAFKSGDAV 2281  
 QY 966 KKNQIGVKT-----TTLTKT---DANG-----DKAINESVN----- 993  
 DB 2282 DVSQGLVDTSGSASAAPHFTTQNTPIISAPGIDTLRPVLQSNNAATSVINEFATSGTGF 2341  
 QY 994 SGDDKALINAKDIANDNTLGEIRNTGTADTAQTQVKKVYKENGDDNDADDTIYV- 1052  
 DB 2342 TGGIENL-----YSNVGEGAVAVNATGASSFLRN-----TTITAGG 2380  
 QY 1053 ---KDAKTN--QVNTLKLKGKMGIDTQNKDGTVEGINTQSGLKA---GNNTTLNNGLS 1105  
 DB 2381 RALDFSTTGAPANTLL-----SIOGNTLK-----SSGSALAASBGOMISATDMSIA 2428  
 QY 1106 IKNITGANOIOVADGVKFAVYN-----NGVAGADGT---TRTRDEIGFAGTNGS 1155  
 DB 2429 IRSPAGNRVGTGTGGGIAFNWVRPDSGAGTVAAGTGMNPGARIQDGLSPANTSGT 2488  
 QY 1156 LDKSPHLSKDG---IN---AGGKKTITNIOSEIIONSN---DAYNGKRIYDKTE 1202  
 DB 2489 FDLGLNLANNNGGVVLTNTGAGTITLNLNSAGVLTNSGAFLDPLTVHMTFASVTS 2548  
 QY 1203 LENKISSTAKTAONSLEHFSVADEQNNFTVSNPYSYDT-----SKTSDVTF-- 1251  
 DB 2549 VNNPANAIA---NGI---IFDGVAGTFVYGTITNTGPFIDAINVTANTGTFQF 2600  
 QY 1252 ---AGENITF---KYNKGVN-----VGIDQTKLT---TRKLYVGNNGK 1289  
 DB 2601 GAVGINNGTGGGIIHVASGTLVNTGLANIDTTSVGLSOGSGTSTFTGGLIDTFTTGT 2660  
 QY 1290 GIV-----IDSONGQHTIT-----GLSNTLANTYNNKGSVRTTEQKRIIKDEKTR 1335  
 DB 2661 GIVGTGTMGTATAGAEVTVSTGGAISLDNIAANVTLD-----STSSGCGTNNVSLIN 2715  
 QY 1336 AASIVDV---LSAGFNLOGGEAVDFVSTYDVNFA-----DGNATTAKVYTD- 1380  
 DB 2716 VSGTVGLGTGLGASGASFVNVGTAVTYGNISSQANNAALVSVSGNA--GSLTFNT 2773  
 QY 1381 ---DTSKTSKVVYDVND-----PTTIEVKDKKGLVKT---TTLTS-TGT 1418  
 DB 2774 GADLATNGTGLQFD-NADGTYNFNGLTTLHGAGADIDILNGSAGTFEFPVTLITSPGT 2832  
 QY 1419 GANKEF-----ALSQATGALVKAS-----DIYA 1442  
 DB 2833 AFNLVGGTASFYSGTITTOGNALVALSISGHSQGYVSFFGPMASANGTGLQDNADGAY 2892  
 QY 1443 HLN---TLSGDIQTAKASQANSAGYVDADGNKVIYDSTDNKYQAOKNDGYDK---T 1495

Db 2893 NFNCTTLNGC---DAGVDILNNGSGSSENFASSTTTTSPCTAPNNAAGSATVYAGTTT 2949  
 QY 1496 KEVAKDKLVAQAQTPDGTLAQMNKSVYINKRQVNDANKKQINEDNMFVKLEKASDNK 1555  
 Db 2950 QGTAGQRIALADTTGG-----AINTTTAVANGLNLTAGTGT 2966  
 QY 1556 TKMAA---VTVGDLNAAVQTPLEFAGDTGTTAKKLGTTLTGKGGOTDNLKLDNNIGVA 1612  
 Db 2987 LIDGAAGVNTVNNASLTGRTGTTILGDANNAT---GTYTFNNVAIDPVAAGTAHAFVVD 3043  
 QY 1613 GTDGF---TYKLAKDLNLSNVAAGCTKIDDKVSPVDSGQAKANTPVLSANGLDLG 1668  
 Db 3044 GDQGPANNDVSATIALNNVDITNPGGNVNIQMS----- 3079  
 QY 1669 KYISVNGKGTQDPAANYQOLNEVNNLLGLG-----NAGNDNAGNOY-----NIADI 1716  
 Db 3080 -----AGSYTFEDNAS-----TTRNDGGLGIVATSNAGGIIAGCAATLTTSVNNAVDL 3128  
 QY 1717 KDPNNGSSNRRTYIKAGTGLGKGNNDTEKLTATGVOGVVDKDG---NANGDLSNVVK 1773  
 Db 3129 TSNP---GATFNFSGGLATTTTSGIGLN-----ATGGGVNTVNTGNTVQATGQILN-WNG 3180  
 QY 1774 TOKDSKALLATYNAGQTYLNNPAEALDRINEGIRFFHYNDGNQEPVVOGRNGID 1833  
 Db 3181 MIISSG---ATFSSLOSSGSIT---GLDAINTLGVSSTGTEGNG--AVTVAGTTGVG 3229  
 QY 1834 S-----SASGKHSVAIFQAKAD-----GEAVVAIGRQTOA 1864  
 Db 3230 SDGIEINSSSAFNFSTXTIATADNGIFLISGANGAVTFSTVAIINTGQAIRLDNNINA 3289  
 QY 1865 GNOISIAIGDNOATGDOS---IAIGTGVN-----VAKHSGAIGDPS 1903  
 Db 3290 VNVN---GSLIGATNDPTGNGVDINGTGNTVTAASYTKTSAGNIVNVSGETACTVAFSG 3346  
 QY 1904 TYKADNSVGNNGNFTATQTDVPGVGNNTTYESN---SVAGLSAISAGTHACTQA 1960  
 Db 3347 SLSATGGFANG-----IDVSGNSGTAVINFNSATKLTITGANTAVMLTNTGATI 3396  
 QY 1961 KKSDD-----TAGT---TTAGATGTGVKFA-----GOTAVGA-----VSYG 1994  
 Db 3397 NFTGGGLVINTTSGTGFNATGATVYVQGTGNTTITTTGTLANTANTTIGSNVTFOSIS 3456  
 QY 1995 ASGAEERRI-----ONVAAGEVSATSTDAVNSQLYKATQC--- 2029  
 Db 3457 ANGAANGVLANNTGSSGLIYTGNGNTALGNSSGGTIONTTGAI---LLTFTKGPSE 3512  
 QY 2030 -----IANATNE-LDHRHONENKANAGISSA-----MAMASMPQAYIFGRS 2070  
 Db 3513 TNINIONANSGIKGTIVTNTFLANSTINNVNTAHTTDDGVTNLSAGSGSTETNLSGV 3572  
 QY 2071 MYTGAIATHNGOGAVAVG-----LSKLSDNQGWFKINGSADTGGHCAAV 2116  
 Db 3573 SITGNVLNNSWQDGISINNAGTISLTTTGNCL--TSASFAAGSAGTAI 3621  
 RESULT 8  
 Q48152 PRELIMINARY: PRT: 1098 AA.  
 AC Q48152:  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HIA.  
 GN HIA.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NONTYPEABLE STRAIN 11:  
 RX MEDLINE-96332658; PubMed-8730864;  
 RA Barenkamp S.J., St Geme J.W. III;

RT "Identification of a second family of high molecular weight adhesion  
 RT proteins expressed by non-typable Haemophilus influenzae."  
 RL Mol. Microbiol. 19:1215-1223(1996).  
 DR EMBL:U38617; AAC3721.1;  
 SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;

Query Match 6.4%; Score 681; DB 2; Length 1098;  
 Best Local Similarity 24.1%; Pred. No. 76-15;  
 Matches 278; Conservative 155; Mismatches 427; Indels 292; Gaps 41;

QY 1146 EIGFAGTNGSL---DKSKPLSKDGINAGGKRTITNIOGSEIADNSDAYTGKRYDLKTE 1202  
 Db 64 DANFNFTNNSIADAEKQVQAYAGKGLNLNENK---ASDKLVEDNTAATVGNLRLKGVN 119  
 QY 1203 LENKISSTAKTAQNSLHESVADQGNFTVSPYSSYDTSKTSVITTAGENGITTKVN 1262  
 Db 120 LSSKNGTRNEKSQOVKXADVLEFGKGVQVT---STSENGKHTTFALAKDLGVATATV 176  
 QY 1263 KGVAVGIDQTKG-LTPPKLTV-GNNNGKIVIDS--ONGQNTI--TGLSNTLAN----- 1311  
 Db 177 SDTLTIGGGAAGATTPPKVNVSTTDGLKFAKADAGANGDTVHLNGISITLDTLVSS 236  
 QY 1312 -VYNDKSVTTEQGIKXEDKTRASIVDVLASGFNQ-----GNGEAVDFVSTY 1362  
 Db 237 PATHIDGDSSTHY-----TRAASIKDVLNAGMINKGVAGSTTGQSENVDFVHTY 287  
 QY 1363 DYNFADGNATTKAVYDDTSKTSKYVYVNDVDTIEVKADKLGV-KITTLTSTGTAN 1421  
 Db 288 DTVEFSLADETTTVTVDSKENGKREVEKIGANTSVIKERDKLFTGKANKETRNKYDAN 347  
 QY 1422 KFLASNOATDALVYKASDIYAHNLNTLSGDIQTPAKGASQ-----ANSSAGYVADGN-- 1472  
 Db 348 -ATEDADECKGLVYAKDYDAVANKTGWRKTTDANGQNDPRTVASGNTVPASSNGTT 405  
 QY 1473 KYIYDSTNKYYQAKMDGYDKTEYAKDKLVAQAQTPDGTLAQMNKSVYINKREYNDAN 1532  
 Db 406 ATVYNGTDG--ITVRYDAKVGDKLIDGKRIA--DTTALTVN----- 444  
 QY 1533 KKGGINEDNAFVYGLEKASDNK---TKNAAVYVGLNNAVQTPLEFAGDTG----- 1582  
 Db 445 -DGKANNPKRGVADVASDEKELVYAKGLVTA--LNSLSWTTTAAEDGTLDMASE 500  
 QY 1583 TAKKIGFTLTKGGQDPTNKLNTONNIGVAGTGFYVYKLAKDLTNNNSV-----NAGGT 1636  
 Db 501 QEVYAGDKVTFFKAKG-----NLKVOEGANFTYSLDALTGLTSLTIGTNGANGART 551  
 QY 1637 KIDDKGVSPVDSGQA--KANTPVLSANGLDLGKAYISNVGKGTQDPAANYQOL----- 1689  
 Db 552 EIMKDGLTTPPACAGANNANTTSYTKKDGISAGQGVKNVYSGLKRFGANPPLTSSAD 611  
 QY 1690 -----NEVRNLLIGLGNAGNDN---ADGNOVNIAD-----IKDPNNGSSSS-- 1726  
 Db 612 NLTKQNDADAKYGLTNDEKGTQKQTPVADNTAATVADTGLGLGVISADKTTGGSTEYH 671  
 QY 1727 ---NRVYIKAGTVLGKGNNDTEKLTATGVOGVVDGDNANGDL--SNVWVKTKQKGSKR 1781  
 Db 672 QVRANAEVK---FKSGGLNVSGKYVNGRREITFEEL---AKGEVYKSNETVYAKETNGKET 725  
 QY 1782 ALLAT---YNAAGQTNVLTNNPAEALDRINEGIRFFHYNDGNQEPVVOGRNGIDSSASG 1838  
 Db 726 SLKVGDKYYSKIDDLITLTPG-----KLKDGK----- 753  
 QY 1839 KHSVAIGFQAKADGEAVALIGRQTOAGNOSIAIGNAQAOTGDOSIAIGTGNVYAGHSGA 1898  
 Db 754 ---TVAAKYQDK-----GKRVASVTDNTEATITPK--GSGYVYVGNQVADA 793  
 QY 1899 IGDPSTYKADNSYSGVNNNO-----FTPATQTDVAGVGNNTVTSNSVAGLSNSAISA 1952  
 Db 794 I-----AKSGFELGLADADAKKRAFDDTKTALSAGTTEIYVNAHDKVRFANGLNTKYSA 846  
 QY 1953 GTHAGTQAK-----KSDGTAG 1968

Db 847 ATVESTDANGKVTTEFTKDVVELPLTOIYNTPDANGKITTUKVVKDQTKWVELNADGTAD 906  
 QY 1969 TTTT-----AGATGTVKGFAGOTAVGAVSVG-----1994  
 Db 907 MKREVTLGAVDSGKKVYVNDGKWKHAKADGTADTKGEVSDKYSTDEKHVYSLDPND 966  
 QY 1995 -ASGERRIQNAAGEVSATSTDAVNSQLYKATOGIAN--ATNELDRIHONENKANA 2050  
 Db 967 QSKGKGVAVDYNVANGDISATSTDAINGSOLYAVAKGVTLNAGOVNMLEGKVKYKRA 1026  
 QY 2051 GISSAAMAASMPQATIPGSMVTGATIHNGCAVAVGSLSDNOQWPKINGSADTGC 2110  
 Db 1027 GTASALAAASQLPQATMPGKSMVAIAGSSYQGGALIGVSRISDNKGVIIIRLSGTTNS 1086  
 QY 2111 HGAAVAGGFHF 2122  
 Db 1087 KTCVAGVGVQW 1098

RESULT 9  
 Q9F2D8 PRELIMINARY; PRT; 1107 AA.  
 AC Q9F2D8: 01-MAR-2001 (Tremblrel, 16, Created)  
 DT 01-MAR-2001 (Tremblrel, 16, last sequence update)  
 DE 01-DEC-2001 (Tremblrel, 19, last annotation update)  
 DE SAPB PROTEIN.  
 GN Salmonella typh.  
 OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;  
 RT "Evolutionary origins of the autotransporter proteins."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ277623; CAC14217.1;  
 SO SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;

Query Match 6.3%; Score 670; DB 2; Length 1107;  
 Best Local Similarity 23.7%; Pred. No. 1.6e-14;  
 Matches 304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;

QY 993 NSGDDKALI--NAKDIADN--LNTLAGELRNKGTG---DTALQTFQVKKVKGNGDDN 1044  
 Db 46 MAGNDTGSINQNTTDIATNTTINSNSVTTLTDALLMDASGTFSSAS--NGSASK 102  
 QY 1045 DADDTI--TVGKDAKTQOVN-----TLKLGKNGLDIQTNRKDGTVFPGINTOSGLKAGNN 1096  
 Db 103 ITLNLAGTILADS--TDAVNSQLEDTNEKYDQNTADITN-----TNSINQNTDIAFNT 156  
 QY 1097 TTLNNGSLSTKNTAGNEQIQVADGVKFAKVNNGVAGIDGTTTRTRDEIGFAGTNGSL 1156  
 Db 157 TSTINN-----LSNSV-----TTLDALLMMAAGTGF 183  
 QY 1157 DKRPHLSKGINAGCKKITTINIOSGEIAONSNAVGTGIYDLKTELENNISSAKTAQN 1216  
 Db 184 SASR-----NGSASKITNLAAGTLAADSTDAVNSQLEDTNEKYDQNTADITNTN 234  
 QY 1217 SLHEFSVADEQNNFTVSNPYSYDTSKTSQDVI-----TF-AGENGITTKV--NKGVVR 1267  
 Db 235 SINQNT--DIATNTSINNLNSVTLTLDALIMDASGTFSSASRSGASKITNLAAGT 292  
 QY 1268 VGIDQTKGLTTPKLTVGNNNGKGIIVDSQNGQNT--ITGLSNTLANVTNDKGSVFTTEQ 1324  
 Db 293 LAADSTDAV-----NGSOLYETNQKVDQNTSAIADINTSITNLSQD-----333  
 QY 1325 GKTIKDEKTRASIVVLSAGFVLQNGEAVDEVSTYDYNFADKANATTAKYVYDDTSK 1384  
 Db 334 -----NLSWN-----ETTNSFSASHGSSSTINKITNVAAGE 363

QY 1385 TSKRVYDVNVDPTTIEVKDKKLGKTTTLTSTGTGANKRFALSQAATGDLVYKASDIYAH 1444  
 Db 364 LSEESTDA--VNSQLEFETNEKYDQNTDIAANTN-----ITQNSTAIENL 408  
 QY 1445 NTLSGDIQTAKKASQANSSAGYVDADGNKYIYDSTQNKYQAKNDGTVDKTRVADKLV 1504  
 Db 409 NTSVSINT-----STGLTD--NALLMDE-DTGAFSAHHGSGTSKITVNA-----451  
 QY 1505 AOAOTPDGTLAONNVKSVINKREQVNDANKKOGINEDNAFVKGLEKASADNKTNAATVYG 1564  
 Db 452 AGALSEDST-----DAVNSQLYETNQK-----VDQNTS-----AIA 483  
 QY 1565 DLNAVAOTPLTEFAGDGTGTTAKLGETTLITKGGQDTNKLTDNNIGVYAGIDGTGVYKAKD 1624  
 Db 484 DIN-----TSTNLGTDLASWDEDEGAFSASHGTSQTKITN-----VAAQ-----ETASD 529  
 QY 1625 LTNINSVNAAGTKIKDKGV--GFVDSGGAKANT--PVLASGLDGGKVISVYNGGK 1679  
 Db 530 ST--DAIN--GSOLYETNMLISQINESISQLAGDSETTYTENGG--TGKYYIRTNNGLE 583  
 QY 1680 DTDA-----ANVOQLNEVENLLGLGNAGNDNAGNOVNADIKKDPNSGSSNRVYKA 1733  
 Db 584 GQDAYATGNATAVGYDAVAVASAGCLALAGNSSSTIEGSA-----LGSGSTSNRAITTG 638  
 QY 1734 -----GTVLGGKGNNDTEKLATGGVGVYDKDGNANGDLSNVWTKQKDGSKALL 1784  
 Db 639 TRETSAISDGVVI--GYNTDBELLGALSIGT--DEESYROIITNV-----ADGSE-----684  
 QY 1785 ATYNAAGQTYLNNPFAELDRINEGIRFEHVADNGQEFVYVGRNIDSSASKHSVAI 1844  
 Db 685 -----AQAAYVYRQLOMALGAVTTPTPKYHANSTEE-----DSLAVGTDSLAM 728  
 QY 1845 GFOAKADGEAAVAIGROTA---GNOSIAIGDNAQATGDSIAIGTNVYVAKHSGAIGD 1901  
 Db 729 GAKTIYAADGIGGLNTLYMADALINGAIGSNARAHAHANSINGNSQTT--RGQOTD 785  
 QY 1902 PSTYKAD-----NSYVGNNN--QFTD--ATQTDVFGV-----NNTVT 1937  
 Db 786 YTAVMQDTPONSVEGFVSQEDGQROIITVNAAGSADPDANVAGOLKVTDAQVSRNTOSIT 845  
 QY 1938 ESN-----SVALGSNSAISAGT--HAGQAKKSDGTA--GTTTGAATGTVKGA 1983  
 Db 846 NLNTQVSNLDTRYNTINENGIDYITGTSYKTFKTNTOGADANAGDASVALIGSSIAAE 905  
 QY 1984 GQZAVG-----AVSAGSAGERRIQNVAA-----2007  
 Db 906 NSVALGTNSVADEANITVSQSSGTQRRITVNAAGVNNTDVAVNQKLKASEAGSVRYETNA 965  
 QY 2008 -----GEVSA--TSTDVANSQLYKATOGIANATN-----ELDHR 2040  
 Db 966 DGSVNSVLMGLDSSGCTTRIGNSAAVNDTDVAVNQKLKRSVSEANTYTDOKKGENNSK 1025  
 QY 2041 IHONENKANNISSAMAMASPOATIPGRSMVTGATIHNGOGAVAVGSLSDNOQWPKING 2100  
 Db 1026 IKGIENKMSGISAMAMAGIPQATYAPGANNSTIAGGTENGESVALIGSVSVSGGMY 1085  
 QY 2101 KINGSADTQGHVGAAGGFHF 2122  
 Db 1086 KLGSTNSQGDYSALIGAGFQW 1107

RESULT 10  
 Q9F285 PRELIMINARY; PRT; 3705 AA.  
 AC Q9F285: 01-MAR-2001 (Tremblrel, 16, Created)  
 DT 01-MAR-2001 (Tremblrel, 16, last sequence update)  
 DE 01-DEC-2001 (Tremblrel, 19, last annotation update)  
 DE YAPB PROTEIN.  
 GN YAPB.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;





Db 2068 VGNITVAVKVIDADATLYVD-----TALLDHVLTENGTLNVAVKYLATTAEDKSGTGVGXXPSGI 2123  
 QY 1569 VAOPTLFEAGDTGTAKLCELTILIKGGQDTKTKLTDNNNGVAVGCTPVTAKLADLTN 128  
 Db 2124 VNLNTTFALISADNAALATL-----KLSDDSTVYTTDRILHGL-----DL 2168  
 QY 1629 NSVAVGATKIDDKGVSEFVSSGAKANTPVLSANGDLGKVIISNMGKTRPD-----AA 1684  
 Db 2169 N-----GGTLI-----FDGSPPOGAN-GVYVYTDIALNSGTLISITGAGWENHEHPYTP 2217  
 QY 1685 NVOOLNEVRN--LLGLNAGNDNADQVNVADIKDPNSGSSSNRYIAG----- 1734  
 Db 2218 NVSLLEDRDIDLQLLIDADNVNMGANDLMLNGITTSISAGQVOSTVQGGYTVANATH 2277  
 QY 1735 ----TVGSGK-----NNDTEKLTG-----GVQVYVCKDGNANGDLSNV----- 1770  
 Db 2278 NYGNTSMGSGGLVYNTLSALELLADGANMLLTBESGLTANRELNAELSGVGLVYVDAQ 2337  
 QY 1771 -WVTKQDGSK-----ALLATVYNAQOTNYLT-----NNPAEADIRINE 1809  
 Db 2338 NGALTTLANGNNRREGTTVTAGELLIGANGAFQOTSLNLIASGASANIINGYRQTVGAVTN 2397  
 QY 1810 QG-----IRFHVADGNOEPVVOGRNIDSSAGSKHVAIGFOAKD 1851  
 Db 2398 TGTVTLNGGELSTDTLINTGMINTVDG-----ILNLENGASSISG-GLTGNGILINIKG 2452  
 QY 1852 GEAVALIROTQAGNOSIATIGDNAQAT--GPDSIATIGTVANAGKHSAGIDPSTVAKDN 1910  
 Db 2453 GDFIISIDNNGLAGQOTNIS--DGASVTLGNGGTTIGTN--LGSSVYDVLDDLVLV--ADNS 2508  
 QY 1911 YSVGNNOPTDATTQDVEGVGNNTVYESNVALGNSAISACTHAGTOAKKSDGTAGTT 1970  
 Db 2509 LA-----NVISGCG--TINTTATVYLLSGNSPFGAHQICGTNELVIGASNL 2553  
 QY 1971 TTAGATGTGAGQTAAGVAVSVA-----SGAERRIONVAAGEVSATSTPAVNSQL 2023  
 Db 2554 GASSAT-----VNLGTLTSLHLINGVSESIANVLG--VAGSTVLIIGAD- 2597  
 QY 2024 YKATGIANTNELHRIHONENKA-----NAGISSAMAMAMPQAYIGRSMVTGIA 2077  
 Db 2598 ---TALTAANSGLFGYALAGNSKLTAYASTNNIGASSVALAG-----TGDTLISLGFN 2648  
 QY 2078 THNGGAVAVGLSKLSDNGQ 2097  
 Db 2649 GTFGNSVTGSGVLQVTDAAE 2668  
 RESULT 11  
 Q92D91 PRELIMINARY; PRT; 2340 AA.  
 AC Q92D91;  
 DT 01-MAY-1999 (T-EMBLrel. 10, Created)  
 DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE CDL SURFACE ANTIGEN (SCA3).  
 GN RP451.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MADRID E;  
 RX MEDLINE=99039489; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Scherf-Litz-Ponten T., Almark U.C.M., Podowski R.M., Naslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria."  
 RL Nature 396:133-140(1998).  
 DR EMBL: AJ235271; CAA14908.1; -  
 DR InterPro: IPR003858; rompa\_rompb.

DR Pfam: PF02708; rompa\_rompb; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;  
 Query Match 6.1%; Score 651.5; DB 16; Length 2340;  
 Best Local Similarity 21.2%; Pred. No. 1.4e-13;  
 Matches 519; Conservative 302; Mismatches 879; Indels 743; Gaps 121;  
 Db 10 NKATGFMAVAE-----YKSHSTG-----GSGCA-----TGQVGSVRLSPARIALAVL 55  
 Db 14 SKILKFLTAALCGTLFNSNATGTIIIPNNGSVLNTDAGLVGV-----FNNGDITOI- 58  
 QY 156 VIGATLNSAVYAGIIGISEADGGKGANARGDKSIAIGDIALGSGSIAI----- 105  
 Db 159 ----VNGREIKISADKANALIGGINLKE-----LPFGGVGVQNVSTIGPLNAGEDLN 109  
 QY 106 --GDNKIYHNS-----NNANIGAK-----ASGNESI 130  
 Db 110 FNGPLKPISSNVVTSITIGVGTKEPSNIDFAGKNATLQINDLTITTKIDNTVAGNNGSI 169  
 QY 131 AIGDVLASGH-----ASIAISDDLYLKKEVYQOISELPIIRGGKALNDIYQLADT 183  
 Db 170 TREGSGIISNHLGYNSLGLINGNGEAKIYAPEANNIT-----I 209  
 QY 184 NLOKVRRTHAQHASAVAGAMSVAGHFSNAPFRTAETGYSLAVGLTAAKASSIAY 243  
 Db 210 NAKNINLTHNNSILITLCOSGITTLKGNINNT--TEIDGGILNLAIDL-----GSSSIIT 262  
 QY 244 GSNQAQIGFAAFAVGSGTQVNIHRCIALGFGSQVYLQKNDVNAANVAYADDPNDPINDR 303  
 Db 263 GD-----IG-----NISLPTINVLGSAT--FNSTLIKATNINLKHNTSTLNLDND----- 307  
 QY 304 YKATFNKATDVEFSIGNSGNSDIRKIIN--VGASADTAVVNAOLKEAVRIANRQIT 361  
 Db 308 -----IIVIGNIKGNN--KDLINKEVHGNTINDKMIIPAPK-----THGLIN 349  
 QY 362 FKGDDSN-----RVKGLGKTLTITGCAQ-----TSA 389  
 Db 350 FNGMNTLNGNINNLNLIKFSGGHGTMLNQGNTKYDNLVFADSVLDSGTISVNGLLDTDC 409  
 QY 390 LUDHNIQVQNDGKLVQLAETL-----ISLKWVTEENLTANKYVGTRLTLTKIG 442  
 Db 410 VFENNSNV--NGSTLIINAKNTISAKLNAATKAKIQIANLTMN-----PSAG 456  
 QY 443 FNDMNGIDESKPYLDKDTGIGHAGQITKTLAGVDDAATYQOLKVNQTAESALOTF 502  
 Db 457 DISDIRIADNTIYTTIDAKN-----GVNLLNNNAKIIIFEGADS--MLALIN--TGYADRTF 509  
 QY 503 TV-KRYDKNGDAND-----SKIITVGNKKNPDGT-QVNTLKLK----- 539  
 Db 510 EIVNNLNGSGNDEYGIKLEAIKVVITIANOSGPYTIQDNTMRKLKELIVGAGDIID 569  
 QY 540 -----GENGVDTTEENGIVTGTGLQNNNGLIYGN 568  
 Db 570 TTFKLSINSTGQITFNKTLDLGAGNIAFGKHTLVVNGVTSITISENNOGILTINS 629  
 QY 569 STLNDGSLVKTNNSKQOVGADGIT-----FTIISNKGCA-----GI----- 608  
 Db 630 G--NITVGITGTELGLKLVNIGADPYTCSANVFASVALTNPSSVILLADGYTLGVEYTH 687  
 QY 609 ENTTRITRDIGFANNSTGLANKPRLPTGINAGKELTIVQSAINPATNGQLDFENR 668  
 Db 688 NNTKGLSLGTC-SNITGIGTNSAALKINIGACA--SNIDSNIYAG----- 732  
 QY 669 LSTANTEKSGSAATIKDLVNSQVPLTFAGDTGPVNTYKGLKELKVGKGTTA--DOLTKN 727  
 Db 733 --STVLTLDQSELTLLNDVAVNSNIITTAGNNGSKLIFTGNGGICIGNICANGAALQEVVEN 791  
 QY 728 ---NIGYVADSTDNSLTYKLAK-----TSLDLAVNNTKTLTASDKVN--VDSGNNTAKL 776  
 Db 792 GTTNIGGTANSQNTVAHSAANVITGLTIGALKKIDGTITIAHGGLVGDIDFPNNKAGKF 851

```

QY 777 QNGD-----LTF-----SKONTGATPATNSKTI-----800
DB 852 ILGGAMIDSVLCNGSVACTLDFIGDGNVTQINGADNANISITINIGDNKNVTIAND 911
QY 801 --VDGLKFTDNGIALDGTYYITRKDYVFAKQDQSDSKSPYLDKDKLKVGEVEITNGI 858
DB 912 IFVDNIHFTNGIIOLOGN--LFTNHIDFGANGCTLEFN-----GNNVTYLNAL 958
QY 859 NAGKATIGLSNLTLD--AANATTG-----HYVQLG-----VDSDEKTRAAISGDVL 904
DB 959 IVMQO--NGILMFTNLKASDDTIGVYKIIINIGQIPQNETIQQNNKLTLLVSVSSSI 1016
QY 905 NAGFNELKNNDKADK---FVSTDYDVEFNG--NATFAKVTYDQKASVAVDVAWDTTH 959
DB 1017 NF-----GDANSQILSPVDTQIKFINNLNETGGI-----ITLDSN--GNNLT 1059
QY 960 LFGADGNKNOIGVYKTTTLTPTDKADKAI-----NFVNSG--DDKALINAKDI 1006
DB 1060 ISGNNGIK--LGSKNELSLINIKKVTYTNDDIQNIHQILINNGALFDQSLTSAKIK 1117
QY 1007 ADNNTLAGELRNTRKGTADALOTFOVKKKEGDDNDN-----DTIVGDAKTNOY 1060
DB 1118 NINGTAVAG-----ATYTLDAINDFNLFTSGMVEKHQDSILELKNSSNTND 1165
QY 1061 NTEKLGKNGELDIQNKDGTFTGINTQSLKAGNNNTLNNGLSIKNTAG--NEOIOVG 1118
DB 1166 HTITL--TSALDPGNNQFGIILITD-----NKLTIDNG--NAVYTLGTANHLK-- 1213
QY 1119 ADGVKFAKVVNV--VGAGIDGTRITRD--EIGFAGTNGSLDKSPHLSKDGINA-- 1170
DB 1214 --QITFASINDGALATKVGINVENVTLINIKDIELEVNANVLFNKNTYATGAININGHD 1271
QY 1171 --GGKRTINQ-----SGELAONSNDAYT-----GKIYDKTELENKISSTAKTAON 1216
DB 1272 FQAGAGVININDIEIDGVSITGNC--VNGTLEFNGSGKVTGLINIMVLQAGAGVSL 1329
QY 1217 SLHEFSVADDEG--NNFT--VSNPSSYDSTSKT-----SD 1247
DB 1330 ASGNVSTTEILOGNNNMLTFRANSHLTTDINKTGGQDLNLYFINGSSVSSIGANAAVGD 1389
QY 1248 VITFAG-----ENGITTKVKGAVRVGID--QTRGLTTPKL----- 1281
DB 1390 IINAGSVNFSNFKSGNIVISDQATQVNNVATADISGNANNGTLKLNHNTPINITS 1449
QY 1282 TVGNNNKGIYDSONQNTT-----GLSNTLANVTNDKGSVR 1320
DB 1450 TILGNNAIG--TIEVANDVTITGLAQONIHFSNATQATLTGLAASQVYNNITTAGNNIH 1508
QY 1321 TTEGKRIKDEKTRASIYDVLS--AGFNLQNG--EAVDFVSRYDVTVPADGNA-- 1372
DB 1509 TLE-----VTDFDIGNDGIIGDANNRLKSIELTNGTYTINSPHYSSITTANNAQGNVKL 1564
QY 1373 --TFAKVTYDTSKTSKVVYDVNDVTI--EYDKKLGVYTTTLTSTGTANKFALSNOA 1429
DB 1565 NIEGITYDLSKIKSLANVOISEDTRIIRDVYSKYLINIDAGKTINFDGDNMNNPRLND 1624
QY 1430 TGDALV-----KASDIYVAHLNITSGD--IOTAKASQAN--SSAGYVAD--GKNYIYDSTD 1480
DB 1625 IPDALIDLVLPRSLSEFNFTTDIKADNLNLFADDTATANKDAVDAHAHIDNGILKEND 1684
QY 1481 NKY--YOAKNDGTVDKTRKVAKDKLVAQAOTPDGTLAOMNVKS-----VINKQ 1527
DB 1685 NAMLTQEKNNNTI--ELASDKFM-----LLOKINKATLTLADNANVLVDNE 1731
QY 1528 VDNANKKOGIENEDNAFVKGLEKASADKTRKAAVTVGLDNLAAVQOTPLTFAGDTGTAKKL 1587
DB 1732 VNTMLNVDIYDLDA--NELKTYTGNVTHNGLLTI-----ITY--FDTLAQKG 1775
QY 1588 GETLITGGGQDTDKLDNNIGVAVAGTGTFTVKLAKDLTNLS-----VNAGGTR 1637
DB 1776 GHILVSGSVNDMSDL--DNLT-----IKIKASHDITNTSDTKHQIYKLETGAIYTP 1826
QY 1638 IDKGVSFVSSGQAKANTPYLSANGL--DLGKVIYSNVG-----GKTRDDTAANVQO 1688

```

```

DB 1827 YPQTKYIIDASEONKVFYKVVADANGVLVLLDTGGR--DDTGRRDTRGRGNTD----- 1878
QY 1689 LNEVRNLGLGN-----AGDNADGNQVNIADIKR--DPN-----SGSSS 1726
DB 1879 --NCCRDNCVGNISNNSSNEAGSSSDKN--YGITVVPPIFDPSPILDTKNNYVASGIAN 1936
QY 1727 --NRTVIACTGVGKGNNDTEKLATGCVGVKDG--NANG-----DLSNWK-- 1773
DB 1937 QLINHVDFGNTTDGKLLNDIGFMSPNRYETTLDBLSKRINVLNGLNBSYGLNIEVEN 1996
QY 1774 -----TOKD--GSKKALLA--TYNAGOTNYLTNNPAEIRINDOGIRFPHV 1817
DB 1997 FLTDIALNNDFTAKIGRLEELSDANTVNGLNTNTLNNK-----INLRL 2045
QY 1818 NDGNPEVVOGRNGIDSSAGSKHSAVIGFQAADGPAVAIGRQIQAQNOGSIADGNAQA 1877
DB 2046 NNNQALIAAGE--DNIVTIGMGSFYGKIKONSKNS--ASGYQNTGGIIGFDYNT-- 2100
QY 1878 TGDOSIATGTVNVAG--KHSGAIGDPTVKADNSYSVG-----NNNOFTDA----- 1922
DB 2101 --DNSTVIGAATYTMADSKYKHKNKNGDRTKAKSNITSYGLYNNLTNNFVEALGYGR 2158
QY 1923 -----TQDVEFGVNNITVYESNSVALGSNSAISACT 1954
DB 2159 NKIKNERKITITTDQIAGKFINFYSYELLGYNLYLSHRT 2201

```

```

RESULT 12
QY 099054 PRELIMINARY: PRT: 6713 AA.
DB 099054
AC 099054
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE EBHA PROTEIN.
GN EBHA OR SAI267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
NCBI_TaxID=158879.
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hoshiyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hiraoka H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL: AF003133; BAB42527.1;
DR InterPro: IPR002988; GA.
DR Pfam: PF01468; GA. 1.
KW Complete proteome.
SQ
SEQUENCE 6713 AA; 722339 MW; AF6EDE226BE488 CRC64;

```

```

Query Match 5.8%; Score 622; DB 16; Length 6713;
Best local similarity 20.5%; Pred. No. 4,1e-12;
Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;
QY 6 KVFENKATGTFFMAVAVYAKSHSTGSGCATGVGVSRILSPFARIALAVLVIGATLNGSA 65
DB 39 ETLINKQGTGP--NTAKTAVDQALNNVNSAKHALNGTQNNNAKQALATVAINASDLNKO 96
QY 66 YAGISIGSEADGGKGANAG-----DKSIAIGDIAQALGSGSIAIGDKKIYHNSNNAN 119
DB 97 KDALK--AQANGAORVSNADVDORNNATELNTAMGOLQHAIDATNTLASSKYVNVADSTKON 155

```

QY 120 -IGAKSNESTALIGDVLASGHASIALGSDLLYKKEVQOISELLPIINGOKALNDIY 178  
 DB 156 AVTTKYNMAEHIISGTPYVTPPEVTEYAAAN-----OVNSAKOJLNDERLIR 202  
 QY 179 QLADTFLQKRYRTHAGHCHSTAVGMSYAKGHFSNAFGRRAECTYSILAVGLTIFARAA 238  
 DB 203 -----VAKONANTAI DALTO-----LMPQAK 225  
 QY 239 SSIAVGSNAQALGFATAVGSGTOVLNRLGALGSGVLOKNDNVNANVAY-----A 293  
 DB 226 LKEQVQANRLDQVSGVQNGOSLNAMKGL-----RDSIANETVYASQNTYDA 275  
 QY 294 PDDNDPIDNRYATEKNGATDVEFSNGSNGDSIRKKIINVGASADTDVAOVAO--LKE 351  
 DB 276 SPNNQ-----STYNSAVSNAKGIIOTNPNPT-----MDTSATQATQOVNNAKNGING 323  
 QY 352 AVRLANROITFK-----GDDSNRVRKGLKTLITIGGAOT 387  
 DB 324 AENLRNAQNTAKONLTLNLSHLLNNOKSAISSQIDRAGHSEVYTAAKMATELMAQMGNLE 383  
 QY 388 SALTENHIGVONG-----DGLK-----VOLAEEL-----TSLKVTENLITAN 426  
 DB 384 QAIHOON--TVKQGVNFTPADAKAKADATYNNVSRRAETIILNKOGANTSKQDVE-----AAI 437  
 QY 427 EKVYVGTBLTDTKIGFTDNMGIDESKPYLDKDTGIIHAGGOKITKLAVGVDDDAATYG 486  
 DB 438 QVNTSAKNAIANDQ-----NVTNAKNAKNAIINLTSIN--NAOK--RDLITKI--DQATVVA 489  
 QY 487 QAKVYV-----QTAESAQTFYKVD-----KNGNDANDSK-----IITYGRN--KRP 528  
 DB 490 GVEAVSNTGTOLNTAMANLQNGINDKANTLASENYHDAOSDKRTAYTOAVTNAEILNKL 549  
 QY 529 DGTQVNTLKLKENGVDVTEFTNGVTFGLNONGNGLVGNSTLNN--DGLSVNTSNKQI 587  
 DB 550 SCSNLD--KAAYENALSOVYTNNAKGAINGHNEQAKSNAATTIINGLOHTTAKOKKLKQ 607  
 QY 588 QVAGDGTFTDISNSKPGA--GIENYTRIT--RDGIGFANNGTSLDANKPRILPTGSINAGG 644  
 DB 608 VOQAQVAVAGVDPYKSSANTLNGAMGLTRNSIQDNTATKNGQNYLDATF----- 655  
 QY 645 KETLVQASINP-----ATNGQOLDPMRLSTANTEKSGSAATIKDLNLSQV--PLTF 696  
 DB 656 RKRNTYNNNAVDSANGVINTSNPMND--ANALQIATQVYSTKNALDGHENLQAKQOTATN 714  
 QY 697 AGDTGPNVTKKLGELIK-----VKGGTTADDTLTKNIGVAVDSTNSLTVKA 745  
 DB 715 AIDGATNMLKAKKADLKAQVYSAQVAVNTSIQOTANEL--NTAMQOLOHIDDEMATQOT 773  
 QY 746 KTLSDLD-----AVNTKTLASDKVYVDSG-----NNTAKLQNGDLTF-- 783  
 DB 774 QKRYRDEQSKTAYDOAVAAKAILNKQGSNSDKAAVDRALQVYTSFKDALNGDAKLAE 833  
 QY 784 -----SKQNTGA-----TPRANSKTIG-----VDGLKFTDNNGILDDGTYITTKKVGFA 828  
 DB 834 AKAARQONLGTINHITMORTALEGOINOATTVDGVNTKYTANLTDAMMSLOGAIN-- 891  
 QY 829 KODGSIDSKPYLDKDLKVGVEVETTINGINAGKAITGLSNTLTDATNATGHTQIGI 888  
 DB 892 DKDALT--RQONTLDADESK-----RRAVYQAVRAEBILKQGTGNGTSKADVUNA 940  
 QY 889 VDSDTKTRAASIGDVLNAGFINLKNNGDAKDFVSTYDVTDFINGNAFTAKVYDGRASKVA 948  
 DB 941 LNAVTRAKAA-----LNGAENLRN-----AKTSATNTINGLBNLQLOKDNLKHQVE 987  
 QY 949 YQVAVNDGTTIHLTGADGKNQ-----IGVYTTLL-----TKT-----DAKGDAKINFS 991  
 DB 988 QAQNV-----VGVNGVNDKGTNTLNTAMGALRTSIQNDNTTKTSQNTLSDASDSKNRYN 1040  
 QY 992 VNSGDKALINAKD-----IADNLNTLAGEIRNTKGTADTALOTFQVKKYKENGDDNDAD 1047  
 DB 1041 TAVNANANGVYINNTNPMNDAMAINDMANOVTTTKAALNGAQLAO----- 1085  
 QY 1048 TTVVSKDAKTQNVNT-----LKLKGKNGLDIQTNKDGTVTFGINTQSGLAKGANNITLNN 1101

DB 1086 -----AKTNATNTINNAODLNOKOKDALKTYQVNAQVSDANNVOH-----TATELNG 1133  
 QY 1102 NGLSIR-----NNGANGQIOVGADGVKFAFVNNGVAG--IDGTRIT----- 1143  
 DB 1134 AMTALKAALADKERTKASGN--YVADQERQAVDSKVTNENIENITNGTNAATLYVNDVN 1190  
 QY 1144 --RDEIGFAGT--NG--SLDKSRPHLSK--DGINAGSKKTYTINOSGEIQAONSNDVATGCK 1195  
 DB 1191 SAQOVNAKATLALGNDNNLRPAKEHANNTIDGL--AQILNNYQKALKKQVQSATTLIDG 1246  
 QY 1156 IYDLKTELEKNTISTAKTAONSLEHESVADE--OGNFTVSNP--YSSYDTSKTS-- 1246  
 DB 1247 VQVTKNS--SOTLNTAMGLRD--SIANEATIKAGQNTYDASPNRNEYDSAVYAAKA 1300  
 QY 1247 -----DYTFEAGEGNTTKYKNGGVVRVGDIDOTGTLTPK-----LVGNNGK 1289  
 DB 1301 IINOTSNEPMEPNTITQ--TSQVTTKEHALNGAONLQAK--TTAKNNLNTSINNAOK 1357  
 QY 1290 GIVIDSQNGQNTIIGLSNTLANTVNDKGSVTEEGKIIIDEDKTR----- 1335  
 DB 1358 DALRNIIDGATTVAGVQETAKATELNNMHSIQNG--INDEPQOTKOTKYLDAPESKKS 1415  
 QY 1336 -----AASIVDLSAGFN-----LQG-----NGE-----AVDFVSTY 1362  
 DB 1416 AYDOAVNAAKAILTKASQGVNDKAAVEQALQVNVSTKTALNGDAKLNEAKAAKQTLGTL 1475  
 QY 1363 DTVNFADGNATTAKVYD-----DTSKTSKVYD-----VANDDTTI-----EV 1401  
 DB 1476 THINNAORNALNDITQATNEGVNTVAKAQAOLDGAMGOLETSIRKDDTTLQSONYQDA 1535  
 QY 1402 KDKRLGVKTTLTSTGTGANKFALSNQATGDALVYKASDIYAHNTLSGDIOTKAGKASQAN 1461  
 DB 1536 DDAKRTAYSQAVNAATILINKTAGCNTPRAD--VERAQQVNTQANTALNGIONLERAKQA 1594  
 QY 1462 -----SSAGVYDADG-----NKVIYDSTDNKY 1483  
 DB 1595 NTAITNASDLNTKQKALKAQVYTSAGRVSAANGVEHTATELNTAMTALKAALADKADTK-- 1653  
 QY 1484 YQAKNDGTVDKTEKVAADKLVAQAQ-----TPROTLAQNMVKS-----VIN--KQOVN--DAN 1532  
 DB 1654 --ASGNVYNAADNKRQAQVDEKTAHEHIVSGTPPTLPSPVNTAANQVYNAKTQLNGNHN 1712  
 QY 1533 KKQGINEDNAFVKLEKFAASDNKTR-----NAAVYVDLNAVAQTPLEFGDGTAKKL 1587  
 DB 1713 LEVAKQANNTAIDBLTSLNGPQAKLKEQVGAFTLLEPNQVTVADMNQT--NTAMKGL 1768  
 QY 1588 GETL-----TIKGQDTF-----NKLTDNNIGVVA-----GTDGFTV 1619  
 DB 1769 RDSIANEATIKAGQNTYDASQKQNDYNNNAVTAAKAIIGQTSPSMIAQEIQAQDVYA 1828  
 QY 1620 KL-----AKDLTNLNSVNA-----GGTKIDD----- 1640  
 DB 1629 KOQALNGOENLRTAQNTAKOHLNGLSDTNAOKDAARQIEGATHEVETVQAONNADALN 1888  
 QY 1641 -----KGVFVDSGQAKANTPVLSANGLDGKVIS-----NYGKG 1678  
 DB 1689 TANTNLKNGIODONTIKQVNFYDAD--EAKRNA--YTNAVYQABOILNKAAQPNPAKOG 1944  
 QY 1679 KDTDAANVOQL-----NEVRNLGLGNAG----- 1702  
 DB 1945 VETALQNVQRAKNEIENGONVAVNAKTTAKNALNLTISINNAQKAALKSQIEGATVAGVA 2004  
 QY 1703 -----NDNAD-----GNOVNTADIKKPDNSGS 1724  
 DB 2005 QVSTMASSELNTAMSNLORGINDEPAITRAQKYTEADRKOTAYVNDVATYAKTLLDTPAGS 2064  
 QY 1725 SSNRTVTK-----AGTVLGKGN-----NDEKILATG 1752  
 DB 2065 NDNKVAVEQALQVRNTAKTALNGDARLENAKNTAKQOALMTSHLTAQKRNLTQOIERGT 2124  
 QY 1753 VQGV-----DKDG-----NANGDLSNVVYKQKDGSKRA 1782

Db 2125 TVAGVGIQANAGTINQANQALROSTASKDATKSSSEDDYQDANADLQNAV-----ND----- 2175  
Qy 1783 LLAITYNAGOTNLTNNPAPALIRINEGIRFHVNDGNOEPVYQGRNGIDSSASGKHSY 1842  
Db 2176 --AVTNAEGIIIS-ATNPEKNNPPTINDOKASQV-----NSAKSALNGDEKL 2217  
Qy 1843 AIGFO--KACAGEAAVAAGROTOAGNOSIATIGNDAQAGDQSIAGNG-----NVVAG 1893  
Db 2218 AAKQOTAKSD-----IGRLITDLNNAO--RTAANAEVDQAPNLAAYTAARKKATSLITAMG 2270  
Qy 1894 KHSGAIGDPSTVAKANSYSGVNNQFTDATQTDVFGVNNITVESNSVALGSN---SAI 1950  
Db 2271 NLKHALAEKONTKRSVNV-----TDADQPRQOAYDTAVTQAEATLTANNANMETQV 2322  
Qy 1951 SAGTHAGTQAK-----KSDGTAGTTTATGATGYKGA 1983  
Db 2323 QAAALNQALNQKNDLNGDNKVAQAKETAKRALASYSNLNNQSTAAVSQIDNATTVADVTA 2382  
Qy 1984 GQ-----TAVGAVSYGASGAEIRIONVAAGEVSATSTDAVNGSOLYKATQGIANTNE 2036  
Db 2383 AONTANELNTAMGOLNGIND-----QNTYKQOVNFTDADQ--GKKDAY--TNAVTNAOGI 2434  
Qy 2037 LDHRHONENKANAGISSAMAMASMPQAYIPGRSMVTG-----GIATH--NGQ--- 2082  
Db 2435 LDKRANGONMTKAO--VEALNQVTTAKNALNGDANVRAQAKSDAKANGLITLILNNQKOD 2492  
Qy 2083 -----GAVAV-GLSKISDNQ 2097  
Db 2493 LTSQIEGATTNGVNSVYKTKAO 2514

RESULT 13  
Qy 0931R6 ID Q931R6 PRELIMINARY; PRT; 6713 AA.  
AC Q931R6; ID Q931R6;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DN HYPOTHEICAL PROTEIN EBHA.  
GN EBHA OR SAVI434.  
OS Staphylococcus aureus (strain Mu50).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacilli/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani U.I., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus".  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003362; BAB57596.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 6713 AA; 722306 MW; BRC0536AC341BF5 CRC64;

Query Match 5.88; Score 619; DB 16; Length 6713;  
Best Local Similarity 20.58; Pred. No. 5.2e-12;  
Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;

Qy 6 KVINKKATGTFMAVAEYAKSHSTGGSCATGQVSVFTLSFARIALAALVIGTIGTSGA 65  
Db 39 ETIINKQGTG--NTAKTAVEQALNNVNSAKHALGTONLNAKQALTAITANGASDLQOKO 96  
Qy 66 YAGIGISEADGSGKANARG-----DKSIAIGDIAQALSOSIADGNKIVHNSNNAN 119  
Db 97 KDAIK--AQANQAGVRSNANDVQRNATELNTAMGOLAHAIIDKTTLTASSKYVANDSTKON 155

Qy 120 LGAKSGNESTIAIGDVLASGHASIAIGSDLLYLKKTETVOQISELLPIIRGOKALNDY 178  
Db 156 AYTKVTAHEHITSGPTVYVTPSEVTAAN-----OVNSAKOELNDEKL--- 202  
Qy 179 QIADTNIQYRPHAGHASTAVGAMSYAKGHFSNAFGTRAEGETYSLAVGLTATARA 238  
Db 203 -----VAKQNNATVAIDALTO-----LNPQAK 225  
Qy 239 SSTANGSNAQIGFAATVAGSTQVNNLNGIALGFGSOYLQKNDVNAANRAY-----A 293  
Db 226 LKEOVQANRLIEDVOSVQINGOSLNNAMKGL-----RDSIANETTYKASQNTYDA 275  
Qy 294 PDONCPIDRRYKATFNKNGTDFVFSIGNSNGNSIRKIIINVGASADDAVAVAO--LKE 351  
Db 276 SPNNQ-----STYNSAVSNAGIITQNTNPT-----MDISATIQATVANNANGLNG 323  
Qy 352 AVLANRQITPK-----GDSNNRVERKGLKTLITGAGT 387  
Db 324 AENLRQAONTAQNLTLSHLTNNOKSAISSQIDRAGHSEVTAKKNNATELNAQMGNE 383  
Qy 388 SALTDHNIQVONG-----DGLK-----VOLAEFL-----TSIKVYTNELTAN 426  
Db 384 QAIHDON--TVKQGVNFTDADKAKRDATYNAVSRAETIILNKTQAGNTSKQVE---AAI 437  
Qy 427 EKYTVKTRILTDKIGFTDMNGIDESKPYLDKDTGHIHAGQKITKLTAGVYDDDAATYG 486  
Db 438 QWVTSKNAALNGDQ-----VYTNKNAKNAKNAALNNLSIN--NAOK--RDLTGTI--DDATYA 489  
Qy 487 QLKRVN-----QTAESALQTFYKKVD-----KNGDANDSK-----IITVGN--NKP 528  
Db 490 GVEAVSNTGTOLTAMANLQNGINDKANLASENHADSDKKTAYQAVTAENILKN 549  
Qy 529 DGTQVNTLKGENGVDVTEETNGVTFGLNONGIIVGNSTLN--DGLSVKNTSNKQI 587  
Db 550 SCSNLD--KAAVENALSOYTNAKGALNGNHNLEQAKSNANTTINGIHLTQAOKDKLQO 607  
Qy 588 QVADADITTTDISNSKPGA--GIENTRIT--RDGIGFANNTSISDANKPRLPTGIMAG 644  
Db 608 VQQAQVAVGADVIVKSSANTLNGMGTLRNSIQDNATKNGVYLATE----- 655  
Qy 645 KELLTQSAINP-----ATNGGOLDPMRLSTANTERKSGSATIKDLYNLQV--PLTF 696  
Db 656 RKNITNNNAVDSANGYINATSNPMND--ANAIQIAQVYSTKALDGTGNTLQAOQTAN 714  
Qy 697 AGDTGPNVTKLGEILK-----VKGKTTADILTKNNIGVADSTDNLSLVKIA 745  
Db 715 AIDGATNLNKAQDALKAQVTSQAVANTSIQQTANEL--NTAMGOLGIDIDENATKQT 773  
Qy 746 KTLSDLD-----AVNTKTLTASDKVTYDSC-----NNTAKLNGDLTF-- 783  
Db 774 QKYRDEQSKRTAYDQVAANAAILNKQGSNSDKRAVRALOQVYSTIDALNGDKLAE 833  
Qy 784 -----SKONTGA-----TPATNSKITG-----VDGLKFPDNNGIALDGTYYTKRGVFA 828  
Db 834 AKAAQONLGLTNHNTMAORTALEGOINQATYDVGVNTYKNTANTLIDGAMNSLQGAIN-- 891  
Qy 829 KQDSLDKSKPYLDKDLKLVGEVEITTINGINAGKAITGLSNTLIDATNATGHTYQIGI 888  
Db 892 DKQATL--RNQNTYLDADSK-----RNAVTAQVTAEGILNKQGTGNTSKADVNA 940  
Qy 889 VDSIDKTRASTIDVYNAGFNKKNGDADPFSTYTPVFINGNATTAAYTVDKASKRYA 948  
Db 941 LNVATRAKAA-----LNGAENLNN-----AKTSATVINGLPLNQLQDQNLKHQVE 987  
Qy 949 YDVNVGTTTHLGGADGNKO-----IGVKTTVL-----TKT-----DAKGDAINFS 991  
Db 988 QAOVY-----VGAVGVDKQNTLNTAMGALRTSIQDNNTIKTSQNTYDADSKNNVN 1040  
Qy 992 YNSGDDKALINAKD-----IADNLTAGEIRNTKGTADTALQTFQYKVKYKENGDDNDAD 1047  
Db 1041 TAVNANNGVYINATNNPMNMAINALMDMANOVNTTKALNGAQNILAQ----- 1085

QY	1048	TIYVKGAKTQOVTE	-----	ELKKGNGSIDTQNKDGVTEGIMQSSIKAGNNTELTNN	1101	
Db	1086	-----	AKTATATPTINNADDLKOKOKDALKTOYVNAAGVSAANNVQH	-----	TATTELG 113333	
QY	1102	NGLSIK	-----	NTAGNEQIQVAGADVCFKAVNNGVYAG	-IDGTTRIT----- 1143	
Db	1134	AMTALKAALADKERTKASGN	-----	YVNADEKQRAYOSKVTNEMNIIINGPRNATLTIVDYN	1190	
QY	1144	--RDEIFAGT--NG--	SLDKSRPHLSK--	DGINAGGKKTINIGSGEIAQNSNDNAVYTGK	1195	
Db	1191	SAASOVNAAKATALMDGNLNLVAKENHANNITDGL	-----	ADLNWOKAKLEQVQSATPTIDG	1246	
QY	1196	IYDTELEKENSISTAKTQANSIHEESVAD	-----	QGNFETYSNF--	YSSYDTSKTS--- 1246	
Db	1247	VQYVKN	-----	SOTLTPAKKGLRD--	SIANEATIKAGQVNTTASPNRREYDSATYATAKA 1300	
QY	1247	-----	DYTFPAGENGITPTVYVNGGVAVGIDQTKGLTTPK	-----	LTVGNNAKG 1289	
Db	1301	IINQTSNPTMEPTNITTA	--TSQYTTTEHNLNGQNLQAOK--	TTAKNNLNLITSIINNAOK	1357	
QY	1290	GIVIDSGONGNTITGLSNTLANVYTNDKGSVPTTEQOK	IIIKDEKTR-----		1335	
Db	1358	DALTRNIDGATTVAGVNOETAKATELNNAMHSLONG	--	INDETOYKQOYKYLDAEBSKKS	1415	
QY	1336	-----	AASIVDVLASGFN-----	LQG-----	NEE-----	ANDFVSTY 1362
Db	1416	AYDOVNAAKAAILTRKASQVNDRAAVEQALQVNVSTRTALNNGDALKNEAKAAKOTIGTL				1475
QY	1363	DTVNEADGNATTAFTAVTD	-----	DTSKTSKVVYD-----	VNVDDTPT-----	EV 1401
Db	1416	THINNAQRNALNDNITTOATVBEGVNTYKKAQOOLDSAMQULETSIRKDDTLQOSQNYODA				1535
QY	1402	KDKKLGKVTTLTSTGTGANKFALSNQATGDALVYKASDIAVHNLTLSGDIOQAKASQAN				1461
Db	1536	DDAKRTAYSAQVANAATILNKKTAGGNTPRAXD	--VERAMAQVAVTOANTALNGIQMLERAKQDA			1594
QY	1462	-----	SSAGVYADAG-----			NKYIYSTDMKY 1483
Db	1355	NTAITNASDLTKOKALKQAQVYSAGRVSAANGVEHTATELTMTALTKRALADADYK				1653
QY	1484	YQAKNDGTVDTKTEVAKDKLVLAQAO	-----	TFPDGTLAQNNAKVS-----	YJN--KEQVN--DAN	1532
Db	1654	ASQNYVNAADANKRQARDEKVTAAHEHIVSGTPTPLTPSDVYTNNAQTQYTNNAKTQOLNGHN				1712
QY	1533	KKGINEDNAPVKGLEKKAASDNKTK	-----	MAVTVGDLNAAVQOTPLTFPAGDTGTTAKKL		1587
Db	1713	LEVAKONANNTAIDGLTSLNGPQAKLEQVGAQTTLPRVQTVRDNAGTLL	-----	NTAMKGL		1768
QY	1588	GETL-----	TIKGGOTD-----	NKLTNNNIGVYA-----		GTGDTV 1619
Db	1769	RDSIANEATTIKAGQVNTYDASQNKQNDYNAAVYNAAKAIIIGOTTSPSMAIOEINQAKDOYTA				1828
QY	1620	KL-----	-----	AKDLNLMSVYNA-----	GGTKIDD-----	1640
Db	1829	KQOALNGQENLRTAQTQAKHNLGLSLDTLNAQDKAQRQIBEATIVYNEVTOQAONNADALN				1888
QY	1641	-----	-----	KGVSFVDSGGAQKANTPYLSANGDILGKVYS-----	NVCGKT	1678
Db	1889	TAMTNLKNGIQDQNTIKQGVNFTFDAD	--EAKRNA--	--YTNAVYQABQILNKKAQGPMTADG		1944
QY	1679	KDTPAANVQOL	-----	NEVRNLGLGNAG-----		1702
Db	1945	VETALQWVORAKNELNGQNVANAKTTAKNALNLTLSINNAQKAALKSOIEGATTVAGVN				2004
QY	1703	-----	-----	NDNAAD-----	GNQVNIADIKKDPNGS	1724
Db	2005	QVSTMASELMTAASNORGINDEALATKAQAOKTTEADRDQOTAYUNDAYTAATTLIDKTAGS				2064
QY	1725	SSNFTVTK-----	AGTVLGSGKN-----			NDTEKLATGG 1752
Db	2065	NDNKVAVEQALQVRNPTAKTALNGDARLNEAKTTAKQOLATWSHLTNNAQKANMLTQIEGT				2124
QY	1753	VQVGV-----	-----	DKDG-----	NANGDISNVWKTQKDGSKKA	1782

Db	2125	TVAGVOGIGQANAGTLNOMAMQLOFSTASKDATRKSSEDYQDANADIQNNAY	ND	2175
QY	1783	CLATVYNAAGOTNLTNTNPAEADIRINEOGIRFEHVNDGNEFVVGGRGIDSSASGKHSV		1842
Db	2176	-ATNNEGIIIS-ATNNEPENNPTIINQKKSQV-----NSASALNGDEKL		2217
QY	1843	AIGFO-AKADGEPAVAIGROTQAGNOSIAGIDNQAOTGDOSTAIGT-----NVVAG		1893
Db	2218	AAVQOTAKSP-----IGRLTDLNNMQ-RTAANAEEVDQAPNLATAAKNKATSLNTAMG		2270
QY	1894	KHSAIGIDDPETVADNSYSVGNNOFTDARTQDYDEVGNNITIVTESNSVALAGSN---SAI		1950
Db	2271	NKHTALEKDNKTRSVNY-----TDDQDPKQAYDTAVTQQAALITNAGSNANENOV		2322
QY	1951	SAGTHAGTOAK-----KSDTQAGTTTGAAGTGVGFA		1983
Db	2323	QALNOLNQAOKNDLNDGNKRYQAOKETAKKALASYSULNNAOSTATTSQIDNATTVADYTA		2382
QY	1864	GQ-----TAVGAVSYVGASGAERRIQONTVAAGEVSATSDAVANGSOLYKATOGIAMATNE		2036
Db	2383	AQNTANELNTAMGQLONGIND-----QNTVYKQOVNPTDQD-GKRDAY--TNAVYNNQGI		2434
QY	2037	LDHITHONEKANNAAGISSAMAMASMPQAVITPGRSWYTG-----GIATH--NGQ---		2082
Db	2435	LDRKANGNMTRKAO--VEAALNQVTTTKMALNDANDVARYKARSADAKANLITLHLNNAOKOD		2492
QY	2083	-----GAVAV-GISKSLDNGQ		2097
Db	2493	LTSQIEGATTVNGVNSYKTKAQ		2514

RESULT	ID	PRELIMINARY;	PRT;	2065 AA.
093BC7	093BC7			
AC	093BC7;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	RSCA.			
GN	RSCA.			
OS	Yersinia enterocolitica.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Yersinia.			
OX	NCBI_TaxID=630;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21437624; PubMed=11553361;			
RA	Nelson K.M., Young G.M., Miller V.L.;			
RT	"Identification of a Locus Involved in Systemic Dissemination of			
RT	Yersinia enterocolitica.";			
RL	Infect. Immun. 69:6201-6208(2001).			
DR	EMBL:AF394827. AAK77860.1; -			
SQ	SEQUENCE 2065 AA; 210024 MW; A30607073682D176 CRC64;			

Query.Match	5.6%;	Score 604;	DB 2;	Length 2065;
Best Local Similarity	20.7%;	Pred. No. 4.4e-12;		
Matches 497;	Conservative 326;	Mismatches 893;	Indels 690;	Gaps 118;

[illegible]

Db 159 LFGKQVAVGVGLIASTSMNOPIISGCTLT-----OKQKLVNOMANLFTTA-GGYI 213  
 QY 227 LAMGLATAKAASSTIAVGSNAQAIPEAPAVGSGTOVNLNREIALGF--GSQV---LQK 280  
 Db 214 ALIGQOVNDQPSGVINTPO-----GKVALASSRYILNIDRGNLLGVYOGQVOTLLQON 268  
 QY 281 DN---DYNAANVAVAPD-----DNQPIDNRKATFKNGATDVFSIGNSGNDI-- 327  
 Db 269 GGLIRADEGVIGLTAQKEMLMNTVIDNTGILQARGLSKKNV-----IYNGSGEYVS 323  
 QY 328 RKTIIV---GAGSADDAVAVOLKEAVL-ANRQITFKGDDSNRY----- 371  
 Db 324 QQGMINVNSQQRG-----NVILIGENIHLVYASKIDARDEGEGKVLVGDWQKKNK 377  
 QY 372 -----EKGKTLTI-----TG-----GAQTSAL 390  
 Db 378 LKKNASVVMKGANIDVSTHOGAGTAIVLSEHYTGFGYDIIHARGSLSGDGGQVETS 437  
 QY 391 TDHN-----IGVONG--DGKLVOL-----AETLT 413  
 Db 438 SQRLQSFGRVDVSAIMGNTGHWLDPAEVNTVSGAESGVVQIGDIPAGYKKAQVFT 497  
 QY 414 SLKMYT-TEMLTANKVYV-GTFLTTDKIGFTN-----DMNGI--DESKPYLDK 459  
 Db 498 PMAVYQIILNTSINAOBKGTNVITTSNGSLTNCRCMNITLQADINKIAGDATTLOA 557  
 QY 460 DTGHAAGOKITLTAAGVDDAATYTGOLKKNQJAEASLOF-FYKRVYD----- 508  
 Db 556 DQNIYSNGNIT-----ATTGKLNLNLISGDSIVDSITLNLNSDVLNCGDILL 606  
 QY 509 KNGDANDSKITTVGKNNKPDGTQVNTLKLKENGVDVTTETNGTVEGLNONGNLVGN 568  
 Db 607 KHANENNAARISIMGR-----YQVGNLTLEBNTGM-----ASYGVNISNANI-- 651  
 QY 569 STLANDGLSVKRTNSNKO-----IQVADGI-----TFPDISNKPQ-----AG 607  
 Db 652 SVAGETRISSGESSNANQOGWRCIDISGVSFAGKGNMSEFTMSNRSMMGFTVATAG 711  
 QY 608 IENTT-----RITRDGIGFANNITSL-----DANKPRLPPT--GINAGKELT-- 648  
 Db 712 DKNITFOANANGSTSGVDFTN--GSLVSKSGNISFEDINGETITQTSGLRIQSGSVGN 769  
 QY 649 NVQSAIN-PATNGQGLDEPMNLSTANTEKSGSAATIKDLYMLSOVLPFASDTGPNYTK 707  
 Db 770 NVNVEINTKGVYGFLLRSHITATAGNISANATTHKGLM-----ISGDTLMSKN 821  
 QY 768 IGEILKVGKGTADDLTKNNIGVADSTNSLYKLAKTISDLDVATKTLTJASDKYTV 767  
 Db 822 ---IKLOGVTTNSTYVADAIKISGNSSSVOVNMAAGNIS-MVAVNKGTEVGS--TI 873  
 QY 768 DSGNNTAKLONDLFEFSKONTGAPATNSKITVYDGL-----KFTDNNOIALDGTIYIKD 823  
 Db 874 SADYINITAAQSGDFENLISYKGFSPFN-VNISADNITLNNISDNDVNTNFFLA-- 929  
 QY 824 KYGFAKQDQSLDKSP-----YLDKDKLVGE--VEITTINGINAGKATGTGLSNTLTD 874  
 Db 930 ---AKGDIKANLSSPNKALYFKNGMSGASQNLNSAVNVSASVEAVTITGLA--SN 982  
 QY 875 ATNATTHVTOIGYVSDTKTRASIGCVLNAGFNLKNNGDAKDFVSTYDTVDFLNGNAT 934  
 Db 983 RMNVTVGR--DISIT--ANNHGMGTGAGIG-----VDYVNEFAK 1017  
 QY 935 TAKVYDQKASVAVADVNDGTTIHLTGADGNKNOIGYKTTLTKDAKGDKAINFSVNS 994  
 Db 1018 NGNFANNNGSK-----SIGIANADIRANEVNLISNT--SKADSVIAREANITTLT 1066  
 QY 995 GDDKALINAKDIADNL-----NTLAGE--IRNTKGTADTALQTFQVKKVKEGDD 1043  
 Db 1067 GN---INANASTFNRGIVIRONTLSAQKELMALATSSASPAI-----IYQGLSDE 1115  
 QY 1044 NDADITTYGK-----DAKTQOVN-TLKLKKNGLDIQTNK--DGIVTFG-- 1084

Db 1116 SRSHLVAGNISILKSGNSKSGNPRSSVNLAVSLTSVGN-IDSNSSVGDGDIYFNNVD 1174  
 QY 1085 INFOGLKAGNNTLNNNGISIKNTAGNEQIOVA-----DQVRAKYNNGVANGI- 1136  
 Db 1175 LNAALG-----NVTVGEALSTATNSVLSIGGNNISKALNGLLIGKALNTSGAGTL 1229  
 QY 1137 -----DGTTRITRDEIGFAGNGSLDKSPHLSKQD-----INAGG 1172  
 Db 1230 FRANGSLSVAGNIALQGETGTGAIRNGIAFYGNATLINIAKDSLSLIGENTGSOITAGG 1289  
 QY 1173 KKITINOSGEIAQNSNDVATGKITYDKTELENKISSTAKTAQNSLHEFSVADQGNFT 1232  
 Db 1290 NGISYLSPTKTLTINNGLSL-----MEGRSTGA-----GINFP 1323  
 QY 1233 VSNPYSSYDITSKTDVITFAGENGITK---VAKGVYPRG-IDQTKGLTTPKLTIVANN 1287  
 Db 1324 IGN-----NTVLANEGEGLIKSGSVAGSVAIGMWNSSGPAT--IEGSTD 1370  
 QY 1288 GKGIIVDSQNGQNTITGLSNTLANVNTDKGSVPTTEQKIIKDEKTRASIYDVLSAGF 1347  
 Db 1371 GSGVHLFSAEHQ-----IKRINVTGSSIOABGLRI-----SGNATITDTALSGK 1414  
 QY 1348 NLQNGEAVDFVSTYDVFVNFADGNATTAKVYDDFTSKSVYVDVNDVDTTIEVKOKKLG 1407  
 Db 1415 SINGSGIKVDSL-----GSGVITHVLDNALNGSTHIGIVE----- 1453  
 QY 1408 VKTITLSTGTGAKKFKALSNQATDALVKASDVIAHLN---TSGDITQTKAGSQAANSSA 1464  
 Db 1454 ---ITSDINGIHOSIIN--GTTDGIYGDIDDKNLHVGTSETDILLTLOGVATTGSGT 1506  
 QY 1465 GYVADGNKVLYDSTDKKYYQAKNDGVDPDKTEVAKKDLVAQAOT-----PDGTLAQ 1517  
 Db 1507 G-IKLDGNLDSN-----SLNGAIVDGLALDISSLTNSGTRNGRASDGIGVQI 1558  
 QY 1518 N--VKSVINKEOVNDANKKOGIMEDNAEYKLEKAASDNKTKNAAYTVGLNAVAQPL 1574  
 Db 1559 NGTSLNGVNGTSSANGVGVQSGSLDSSHVTGJASGSGKVYD-----AETVL 1606  
 QY 1575 TFAADTGTAKKIGETLT-----IKGGQTD-INKLTLDNNIGVAGTDEFTYKLA 1623  
 Db 1607 NNTMLKSTSDIGVEITANLSGYGSSVQGDVDTGVBIRDRNTLLVGG----- 1656  
 QY 1624 DLTNLNSVNAAGTKIDDKGVSFVDSGQAKANTPVSAN-----GLDLAGKIVSNYKGT 1678  
 Db 1657 DTDDLVIYINGATDQDQGRQIOL--NGNNTLMDTTLTAGNASDGTGIDIDP-LTKNGNST 1712  
 QY 1679 KD---TDAANVQOLNEVRNLLGAGNAGNDANQVNIADIKDP---NSGSSSNRTVIK 1732  
 Db 1713 VDKRATDGDGQVLNGAISG--GYVNGTSDPGSGLKVD--GDSELDNATLNGNSPDGKIEI 1769  
 QY 1733 AGTVLGGKGNNDTEKLATGCVQVGVYDKDGNANGDLSNWWKTKOKDSKRALA-TYNAAG 1791  
 Db 1770 VANLTGNGHSAVHEETAEGS--GVDIGONA-----TLTGGTNDLLAVTGNASG 1816  
 QY 1792 QTNLTNNPAAIDRINEQGIREFHYVNDGNOEPVQNRGIDSSASGKHAVALGPOAKAD 1851  
 Db 1817 D-----VGTGYO---LDCNNTLMDTTLTAGNA-----SD 1841  
 QY 1852 GEAAVAIGROTQAGNOSIAIGDNAQATGDSIALGTGNVAVAGKHSAGIGDPSTYK----- 1906  
 Db 1842 GHGIEYVGPVSSGNFTI--NGNTVGD-----GCVHIDGPMGGLVNGNSANNHGIY 1892  
 QY 1907 ---ADNSYSVGNNOFTTATQTDVF--GVGNNTIYTESNSVALGSNSAISAGTIAGQ 1959  
 Db 1893 LNDYAAINDITLGNAGSGKSLMEIYLPKNIGVNTI--NGKPIDKNSVCGRTKSGSTL 1949  
 QY 1960 AKKSDGAGTATTAGATGYKAFAGQTAAGVAVSGAERRIONTAAAGSATSDAVN 2019  
 Db 1950 ISPPASTSAPTPSPVPTPISL-----LTPILISG-----ENTISQITQPOEISKH 1995  
 QY 2020 GSOYLKATQCIANATNELDHRIT-----HONENKANANGISSAMAMASMPQAVTIGRSMT 2073  
 Db 1996 GSLMKRNOILSS---LDEQIILPVLVTESEERDIANIS--VVVCIPE---GETTES 2044

OY 2074 GC1ATF 2079  
Db 2045 GPCETH 2050

RESULT 15  
ID Q98LN6 PRELIMINARY; PRT: 3659 AA.  
AC Q98LN6;  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE ML0950 PROTEIN.  
GN ML0950.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099; PubMed=11214968;  
RX MEDLINE=21082930; Pubmed=11214968;  
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Matenabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
Tateuchi C., Yamada M., Tabata S.;  
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RT DNA Res. 7:331-338(2000).  
DR EMBL, AP002996; BAB48427.1; -  
KW Complete proteome.  
SQ SEQUENCE 3659 AA; 337498 MW; 7B7BC7EC5C0F3816 CRC64;

Query Match 5.6%; Score 600.5; DB 16; Length 3659;  
Best Local Similarity 21.5%; Pred. No. 1,1e-11;  
Matches 524; Conservative 234; Mismatches 855; Indels 821; Gaps 115;

OY 29 GGGSCATGV---GSVRLSFARIALAVLIGATLNSAY-----AGIGIEADGG 77  
Db 1101 GGGSAATVAVTSTGDTITTYGDOAIGLAQSVGGGNGSTVSLALAKAGIGV--ALGG 1158  
OY 78 KGSANRG-DKS-IAIGDIAQALG-----SOSIAGDNKIVNNSNNNTI 120  
Db 1159 KGSAAAGNDLVVISTGISTGAGTSGRGCTGAAGIILQSVGG-----GGNGCF 1210  
OY 121 GAKASGESIAIGDVLASGHASIAIGSDLYLKETVQOISELLPIIRGKALNDIYOL 180  
Db 1211 AGTLGGGKSTAVG---VAFGGSGAGGSADI-VKVTSTGNIS-----1248  
OY 181 ADTNLOKRTTAAQGHASTAVGAMSTAKGHFSNAFETRTA-----ECTYSLAVG--- 230  
Db 1249 ---TNPDN-----SSGIIAQSTIGGGGNGCFVAVTGAIGDPDAATASVAIGKGG 1296  
OY 231 -----LEATAKAASSIAVGSNMAQIGFAFPAVGGQVULNKGIALGFS 275  
Db 1297 VGGVGDVTVTSTGTTTGTGKFSNGILAQSIGGGGNGGFAVAGSATTTG-MAGIGVAGG 1355  
OY 276 QVLOKNDVNAANVRVAPDNDPINDRYKATFKNKATDVFSIGNSNGDSIRKKIIN-- 333  
Db 1356 -----TGAT-----GSASG-----KVIIVNSY 1371  
OY 334 --VAGASADTDVAVNAOLKEAVRLANKQTTFKDDSNNRVEKGLKTLITTTGAQTSALT 391  
Db 1372 SLVGAAG--QVPLAPASNIY-----SLVTQDSSGIFAQSVG-----GGGGSGGGA 1417  
OY 392 DHNIGVQNDGLKQVLAELTILSLKWTTEMLTANKEVTVTKRRLTDTKIGFTNDMGND 451  
Db 1418 G-SLGVGLGGGLGVSIGGAGS-----GSTAETVTVTSYNNIITGGK 1459  
OY 452 ESKPYLDKDTGTHAGGOKITKLTAGVVDADATYGLKKVNGTAESALQTFTVKRVKNG 511

Db 1460 DSEFMAQSVGGGGGNGGFAALAGSKDMAASV-----AVGSG 1498  
OY 512 NDANDSKILTVGK--NNKPDGTQVNTLKLK-----GENG-----VDVTEETNGTVTGLN 559  
Db 1499 SSGGDAVAVTVISHGNIETDDGSHAFPAQSVGGGGNGGAGIAGLITGEGSLAVG 1558  
OY 560 QNNGLTVNGSTLNNNDGLSVKNTNSNKOIOVGADITFTDISNKPAGIETNTRITRDCI 619  
Db 1559 GS-----GONGAKAVTVYSTGDLRTGKADGILQASIGGGGNGGFGSALAIT---- 1609  
OY 620 GPANNTGSLDANKPRLTPTGNGGKELTNVQSAINPAINGGOLDMMNLSTAN--TEKS 677  
Db 1610 ---TGAGAI-----GVVGGG-----GKGGNADTVYTSTGNIYTLKN 1645  
OY 678 GSAATIDLYNLSOVPLTFAGDTPVNT-----KRLGEILKY-----KGGKTADDLTKN 727  
Db 1646 GNGSI-----LAQSVGGGGNGGAATVTAAGSQKMAAASVGGGAKGG--TSQLVTVN 1697  
OY 728 NIGVADSTNSLTVKLAKTLISDLDAVNTKTLTASDKVTVDS-----GNNTAKLNG 779  
Db 1698 NIGTI-DTGDKNAGIILQSIG--GGGGTGGAISEMVDGAGAAVSGNGGTGDDG 1754  
OY 780 DLTRSKONTATPATNSKTTGVGDKFTDNNGIALDGTTYITDKVGFAPAKODSLDKSP 839  
Db 1755 GRVIVNSNIGTTLANNATLHTVG--GDSNG-----TFQSVGGGGGGGGSF-- 1799  
OY 840 YLDKDKLVGFEVETTN-----GINAGKRA-----ITGLSNTLTDATNATTGHV 883  
Db 1800 -----GAVSTGONAKAALGVSVSGSGAGSGDGKTIVTVSYNDILTG--NGANGIL 1849  
OY 884 TO-----IGIVDSTDKTRAASIGDVLNAGFNKLKNNDKADVSTY 923  
Db 1850 AQSVDGGGGNGGFFAATIKIGSLGKPNKNGAVSVSLGGGAGSG-----GDAEKV--TV 1901  
OY 924 DTVPFI-----NGNATTAATVYDGKASKAV 949  
Db 1902 DSTGIITSGKAFVLAQSVGGGGTGLSVAALNLGEGGNOITRAVGBEGGGGTG 1961  
OY 950 DVNV--DQTTIHLTGADGNKNOIGVTKTTLTKTDKADKAINFSVNSGDKKALINAKDIA 1007  
Db 1962 EVLLTRHGSTI-TTG-----DQSVGLFPAQSVGGSGGNGMAIS--GVINGTD-----AKTIS 2010  
OY 1008 DNLMTLAGEIRN-TKPADT--ALQTF-----QVKVKEGNDDDNDADDTITVGDART 1057  
Db 2011 ASVGFAGGAGTNASKVYVDNTGAISTYGVESDAIOAQSTIGGGGNGGMAVSAVIGSLGTG 2070  
OY 1058 NOVNTLKLKGNGLDIDTNKDGTVTFGINQSGIKAGNNTLNNNG-----1103  
Db 2071 TNFAGVTVGFGGAGFAGDVETVNHGILLDTGLLAAGOTTINDGAYGIFAQSVGGGG 2130  
OY 1104 -----LSIK-NTAG--NEOIOVGADGVKPAVNGVVG-----AGID-----G 1138  
Db 2131 SGGNAITVGLKGNNGAQTVNVSAVG-----GSAGNGTGNQVNTIROYGGEITNGMG 2184  
OY 1139 TTRTRDEIGFAGNIGSLDKS-----KPHLSKDGINAGKKTITNQS--GELAQMSND 1189  
Db 2185 AFGIILAQSIGGGGCTGGRANSISLQLAGKCTLPVCEPAGKPMWNLQATVGGAGGTGND 2244  
OY 1190 AVT-----GKTIYDKLELNKISSTAKTQNSLHESVADQ--GNPFTVSHPYSYSY-- 1241  
Db 2245 AATVDVGN--YDFTTDDGK--SSGIVAQSTIGGGGIGDGVAVGTGLLAIPYVPVPTL 2300  
OY 1242 -----TSKTSPIVTFP-----GENGITTKVNGGVVRVIGIDQTLGLTTPKLTIVGNN 1287  
Db 2301 LKPLGTSSLPSTSGVVALAGNGAGGNGGTVIVANEGVITTHGKSPDIIHQASIGGGGD 2360  
OY 1288 GKGIIVDSQNGNTITGL-----SNTLANYTNDKSVRTTEQKII-----1338  
Db 2361 GG-----DGFAGALGVILGGGKAAGNGSVTVTNGKANLSQAGAAIITQGTSPTPPAE 2416  
OY 1329 ---KDEKTRASTVVDLSNG-----FMLOGNGEYVDFSVYTVNPAADNAT 1373



Db 2417 GAPSADPEQYSAGIFPAQSVGGGTGGAGALLSLGSGKA--GGTGHVTVNNYGGILL 2475  
QY 1374 TAKTYDDTSKTSKVYDVVDVDTTIEVKDKLGVTTTTLSTGTANKFALSNOATGDA 1433  
Db 2476 HA-----DDSVGIFAQSIGGGGAGGSLGISAIAG-- 2506  
QY 1434 LVKASDIVAHLNLTLSGDIQTAKGASOANSAGYVDADGNKVYDSTDNKYQAKNDGTVD 1493  
Db 2507 -----GSGGASGSGGVYVNNMT----- 2526  
QY 1494 KTEVAKDKLVAAQTPDGTLLAOMNVKSVINKQVNDANKKQGINEDNAFVKLEKASD 1553  
Db 2527 ETHGI--DSYAIOAQSVYG----- 2547  
QY 1554 NKTNAAVTVYDNLAAVAQPIFPAGDGTAKKLGFTLTIKGGQDTPNKLTDNNIGVAG 1613  
Db 2548 GGGKNGCIT--GSIPL--ISIGAGSSG--IGIVNY--TNNNSL--HTYG--AG 2592  
QY 1614 TDGFTVLAADLTNLSVNAAGTPIIDKGVSPVSSGOA--KANTPVLANGLDLGGKV- 1670  
Db 2593 ADGI-----NAQSIGGGGS--GGRAIGFAVGGKGGDVGNGTTAGSNG--GGGAVT 2640  
QY 1671 ISVVGKTDTPAANQOL-----NEVRLLGLGNAGNDADGNQVN 1712  
Db 2641 VNNNANGITVEIGHGIFPAQSVGAGGSGGAFGVSIYPVAIGVGGGSGSDGCVYT 2700  
QY 1713 IADIKDPNSG--SSNRIVKACTVLGKGNNDTEKLT--GGVYGVDKDGNANG 1765  
Db 2701 V-----NNHGDITTVSASSVALFAESIGGGGTGAMSVASPAFATFEGIGDGGANG 2754  
QY 1766 DLSNVVKTOKDSSKALLATYNAAGOTNYLTNPALIDRIINEGIRPFHNDGNQEPV 1825  
Db 2755 KGGNVNVTNFSOG-----IHTGGFEST-----GMAQSVGGG-- 2788  
QY 1826 VQGRNIDSSASGK--SVAIGFOAKADGEAAVAIGROTAQNSIAIGDNAQTGDOS 1882  
Db 2789 --GAGGASYSVSGPPGLAIALGKCAAGDGIY--TVINNGAM-----QLDGDNS 2836  
QY 1883 IAGTGNVAVAGKHSG-----AIGDPTVKADNSYSVGNNOFTDATOTDVFVGNNITV 1936  
Db 2837 VALFAQSVGGGSGGTATAAALGVPTIGD-----TGAT-----GKGGDVTV 2880  
QY 1937 TESNSVALGSNSAISAAGTHAGTQAKSDGTACTTTAGATGVKGFAGOTAVGAV----- 1991  
Db 2881 TMTGQIRLTGNG--SVGIFA-----QSVGGGGVYTAGTGIQAVAGSGGNGVYTTNS 2933  
QY 1992 -----SYGASGAERIRIONVAAGEVSATSDAVNGSOLYKATOGIANTNELDHRI 2041  
Db 2934 NVAMLITGDNSVYFG-----OSIGGG--GGVGGFSGNYL-----GLDQVQS----- 2973  
QY 2042 HONENKANAGISSAMAMASMPQAYI--PGRSMVTGGIATHNGGAVAV----- 2087  
Db 2974 -----TSSKMAA--POGFMSAGGGGTGATFTOTADLAVIGKNSFALMEQSA 3021  
QY 2088 -GLSKLSDNGQWVFKINGASADTQGHVGAAGAF 2120  
Db 3022 GGTGNTSDNGDINVTIASGVTTIIGSGAGGIGY 3055

Search completed: July 30, 2002, 15:38:43  
Job time: 408 sec

